Technische Universität München

Fakultät für Sport- und Gesundheitswissenschaften

Lehrstuhl für Trainingswissenschaft und Sportinformatik

Network Analysis in Team Sports

Florian Korte

Vollständiger Abdruck der von der Fakultät für Sport- und Gesundheitswissenschaften der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Philosophie (Dr. phil.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. Filip Mess

Prüfer der Dissertation:

1. Prof. Dr. Martin Lames
2. Prof. Jaime Sampaio Ph.D.

Die Dissertation wurde am 03.09.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Sport- und Gesundheitswissenschaften am 04.12.2019 angenommen.

In Erinnerung an
Onno Schröder

Contents

List of Figures I

List of Tables I

List of Publications II

Summary III

1 Introduction 1

1.1 Performance Analysis in Team Sports . . . 1

1.2 Social Network Analysis in Team Sports . . . 3

1.3 Critique of Current Research . . . 7

1.4 Aims of Dissertation . . . 11

1.5 Outline of Dissertation . . . 13

2 Methods 14

2.1 Historical Background . . . 14

2.2 Network Modeling . . . 17

2.3 Network Metrics . . . 23

2.4 Related Work . . . 34

3 Studies 39

3.1 Study 1 - Characterizing different team sports using network analysis . 39

3.2 Study 2 - Passing network analysis of positional attack formations in handball . . . 41

3.3 Study 3 - Play-by-play network analysis in football . . . 43

4 Discussion **45**

4.1 Theoretical Impact 46

4.2 Practical Impact 49

4.3 Limitations of Method 50

5 Outlook **52**

6 Bibliography **53**

A Appendix **63**

A.1 Study 1 - Permission for the inclusion in dissertation 63

A.2 Study 1 - Original article 64

A.3 Study 2 - Permission for the inclusion in dissertation 75

A.4 Study 2 - Original article 76

A.5 Study 3 - Permission for the inclusion in dissertation 94

A.6 Study 3 - Original article 95

List of Figures

1 Network of 15th century Florentine marriages by Padgett and Ansell (1993) . . . 5

2 Illustration of two passing networks as modeled by Pena and Touchette (2012) . . . 6

3 Exemplary passing network to illustrate flaws in current betweenness application . . . 10

4 Sociogram of friendships in a NYC elementary school by Moreno (1934) 15

5 Network visualization of table tennis matches played in Italy by Lai et al. (2018) . . . 16

6 Exemplary directed and weighted networks . . . 19

7 Exemplary network representation in matrix form . . . 20

8 Exemplary walk, path and geodesic . . . 21

9 Example of a shortest path in a weighted network . . . 22

10 Exemplary of shortest path in a passing network . . . 23

11 Exemplary passing network for degree measurement . . . 25

12 Exemplary passing network for closeness measurement . . . 27

13 Exemplary passing networks for density measurement . . . 31

14 Exemplary passing networks for centralization measurement . . . 32

15 Exemplary passing networks for heterogeneity measurement . . . 33

16 Illustration of two passing networks as modeled by Duch et al. (2010) . 38

17 Overview on studies by level of analysis and modeling of dynamics . . . 45

List of Tables

1 Classification of studies in line with aims of dissertation . . . 13

List of Publications

Korte, F., & Lames, M. (in press). Passing network analysis of positional attack formations in handball. Journal of Human Kinetics.

Korte, F., & Lames, M. (2018). Characterizing different team sports using network analysis. Current Issues in Sport Science, 3:005.

Korte F., Link D., Groll J., & Lames M. (2019). Play-by-Play Network Analysis in Football. Front. Psychol. 10:1738.

Summary

Team sports can be described as complex dynamic systems based on the frequent interaction of various players. While the interaction can take on many forms, e.g. communication on the field, it is frequently studied by analyzing passes between players. Recently, social network analysis has been introduced to the study of sports dynamics in order to quantify the involvement of individual players in the interplay and to characterize the organizational processes used by teams.

Although team sports, in general, rely on the interactive coordination between team members, the method is almost exclusively applied to football thus far. Moreover, as the current analysis is based on one snapshot of the passing network across a match, it does not account for varying situational contexts that influence interaction patterns. Third, current network approaches disregard the dynamics of interactions in team sports.

Hence, the first study of this dissertation project extends the current research beyond the sport of football. The second study implements a play-level approach in handball to break down the analysis to individual plays aiming at a comparison of interplay structures in different situational contexts. The third study introduces a novel network metric to football that regards the sequential order of interplay and, thus, pioneers the modeling of dynamics in interaction networks.

Overall, this dissertation project contributes to a more accurate modeling of team performance under various situational contexts. This, in turn, can be used for practical match analysis in team sports beyond football.

1 Introduction

"One man can be a crucial ingredient on a team, but one man cannot make a team."

- Kareem Abdul-Jabbar

In the 100m discipline, the skills and abilities of one man (or woman) structure and shape the performance prerequisites for competition. Thus, the identification of performance-relevant traits such as reaction time, speed, sprint endurance or power is rather straightforward (Letzelter & Letzelter, 1982). In contrast, this process is more complex in team sports like football because performance is shaped by the collective actions of many individuals (Hohmann & Brack, 1983). Here, the identification of relevant factors for success is impeded by the non-linear properties of team performance (Sampaio & Leite, 2013). In other words, a sports team is more than just the sum of its parts and, consequently, team performance cannot just be deduced from the skills and abilities of individual players. Admittedly, in a 100m sprint, outstanding capabilities of an athlete might not necessarily result in a winning performance, either. Here, success is also subject to non-linearity between performance prerequisites and competition outcomes as the link might be disturbed by the athlete's form on the day or other external influences (Hohmann et al., 2002). However, the interaction process between players in team sports adds an additional layer of complexity to the understanding of the relevant factors for the observed game performance (Lames & McGarry, 2007). The reason why it is crucial to understand the structure of performance in any sport is that training is shaped by insights generated from past performances. This follows from the fact that players rely on feedback to improve their performance (Maslovat & Franks, 2008). Thus, advances in the analysis of team performance benefit training and coaching (McGarry, 2009). As a necessary precondition, this requires a profound understanding of team sports as such.

1.1 Performance Analysis in Team Sports

In general, team sports can be described as complex dynamic systems characterized by the frequent interaction of many individual parts with the common goal of winning (Grehaigne et al., 1997; McGarry et al., 2002; Davids et al., 2013). In particular, the aim of a team is to score more points or goals than the opponent. According to Mateus (2005), this requires a balance between offensive in defensive actions. Teams must push through their own strategy while preventing the implementation of the opponent's strategy (Davids et al., 2005). This results in the emergence of complex interactions, which is not referring to inter-limb or intra-limb coordination as described by Glazier (2010), but the inter-personal coordination between and within teams (Mateus, 2005). However, the interactions of players is not just shaped by the instructions of coaches or captains to push through the strategy of the own team (Passos et al., 2016). In fact, coordination in team sports is believed to emerge under a set of constraints (Araujo et al., 2006). They are described as factors that shape, limit and reduce the available configurations of dynamic systems such as sports teams. Following Newell (1986), the complex interactions of teams emerge under three different types of constraints: performer, environmental and task constraints. In team sports, performer constraints refer to individual characteristics of players such as motivation, skill level or physical appearance. They can be summarized as the resources to fulfill game-relevant tasks such as the execution of a difficult pass (Passos et al., 2016). Environmental constraints are mostly of physical nature such as wind conditions, temperature or constant factors like gravity. Task constraints include the goals and rules of the underlying sport as well as the boundaries and markings of the playing field. They also encompass informational constraints such as the situational perception of a player in a particular moment of the game which is shaped by the individually available spatio-temporal information (Bastin et al., 2006). According to Araujo et al. (2013), the resulting perception influences the decision-making during the goal-oriented behavior of a team.

There have been several approaches to capture, model and describe team performance emerging from the complex interactions under constraints. Traditionally, a notational approach was taken to especially analyze tactical aspects of performance (Hughes & Franks, 2007). The resulting performance indicators count the frequency of critical events to describe all or some aspects of performance (Hughes & Bartlett, 2002). Following a differential approach, they often aim at distinguishing between successful and unsuccessful performance (Hughes & Franks, 2007). For invasion games like football, events such as the number of shots on goal, corners, scored goals, won tackles or the passing distribution of a team are tracked. While performance indicators, resulting from a traditional notational approach, present comprehensible and informative insights, they often struggle to capture and represent the complex emerging interactions between players and teams (Ramos et al., 2018).

Thus, more sophisticated approaches were recently introduced to the performance analysis of sports teams. These novel methods aim at capturing the complex interactions in team sports (Passos et al., 2016). This is induced by technological progress which offers more accurate and precise data, especially on the position of players and the ball. The resulting approaches predominantly focus on the patterns of coordinated movement between and within teams to characterize performance in team sports. This is in line with Sampaio and Leite (2013) who state that performance indicators may just describe the process of team performance and do not necessarily have to be associated directly to performance outcomes.

There is a body of studies that calculate centroids to track the collective movement of a team (Frencken et al., 2011; Sampaio & Maçãs, 2012). Centroids use positional data to determine the mean position of a team on the pitch. The collective variable reduces the data overflow by describing the movement behavior of a team as a whole. That way, the relative collective movement between opposing teams can be analyzed. Frencken et al. (2011) found that the probability of scoring in football significantly increased once the attacking centroid overtook the defending centroid from a longitudinal perspective. This underlines the occurrence of coordinated movement between teams.

Other studies focus on a relative phase analysis in team sports to detect synchronized behavior between and within teams (McGarry et al., 2002; Bourbousson et al., 2010). Within teams, relative phase analysis is able to calculate the synchrony in movement between players of a team, both in the lateral and longitudinal direction. That way, the coordination properties of defensive lines can be assessed to detect instabilities that the opponent team may take advantage of (Travassos et al., 2011). Similar to this approach, Duarte et al. (2013) apply a cluster phase method to show that the collective movement of a football team is more synchronized alongside the pitch than in a lateral direction. This shows that there is also coordinated movement within teams.

Following the idea of collective action, Duarte et al. (2012) describe sport teams as superorganisms. The concept, which is adapted from sociobiology and traditionally applied to animal collectives, models collaborating players as a single social unit that jointly carries out a task through the division of labor.

1.2 Social Network Analysis in Team Sports

In general, it has become practice to introduce methods from alien fields of research to the performance analysis of sport teams. This also includes social network analysis (SNA), often referred to as just network analysis, which has recently gained in popularity to characterize performance by analyzing interactions between players of a team. The method originates from empirical sociology and was developed to understand and explain interpersonal relations such as friendships or professional work relationships (e.g. Moreno, 1934; Roethlisberger & Dickson, 1939). Wasserman and Faust (1994) describe SNA as a distinct research perspective that models and analyzes complex systems consisting of individual parts which, themselves, are interrelated or interacting in a certain way. The internet is a network linked by data connections, while societies are shaped by people connected through different kinds of relationships such as family ties or friendships. Similarly, harbors are linked by trade routes creating logistics networks. One can study the individual parts, e.g. how a person feels in a social network, or the nature of an interconnection such as the dynamics of a friendship. However, network analysis does not follow an isolated view but focuses on the patterns of interactions between the individual parts, which is believed to determine the way a system behaves as a whole (Newman, 2018). While the structure of the internet affects the way that data is transmitted through the network, the setup of a logistics network determines the shipping route of a package.

In most studies that apply SNA, networks are represented by graphs or matrices (Ribeiro et al., 2017). Graphs model networks as sets of nodes connected by edges. While nodes represent the individual parts, edges capture the links between them (Newman, 2018). That way, graphs can highlight the strategic position of nodes within the network or illustrate the general structure of the interrelations in the system. In contrast, matrices formally and numerically capture the interactions between nodes as a basis to compute so called network metrics. They quantify the individual importance of each node in the overall network, often referred to as the centrality of a node (Bavelas, 1950), as well as the structural properties of the network, e.g. the level of interconnectivity.[1]

To provide an illustrative example, Figure 1 depicts the marriage network of the most influential families in Florence in the 15th century in a graph as modeled by Padgett and Ansell (1993). Their study aimed at an explanation for the strong influence of the Medici family at that time which the authors believed to find in the strategic arrangement of marriages. Their network model consists of nodes, each representing a certain family, and ties that represent marriages between members of these families. As the Medicis were neither among the wealthiest families nor in possession of any formal political power, the authors attribute the influence of the family to their social relationships. When analyzing the importance of the Medici family, the first observation is that they entered into the highest number of marriages as visualized by the number of their direct links. However, they did not distance the second highest family by much. By contrast, their advantageous strategic position within the network is what shaped their influence. Assuming that marriage was the most important form of

[1]A more formal introduction to network analysis and network metrics is provided in chapter 2.

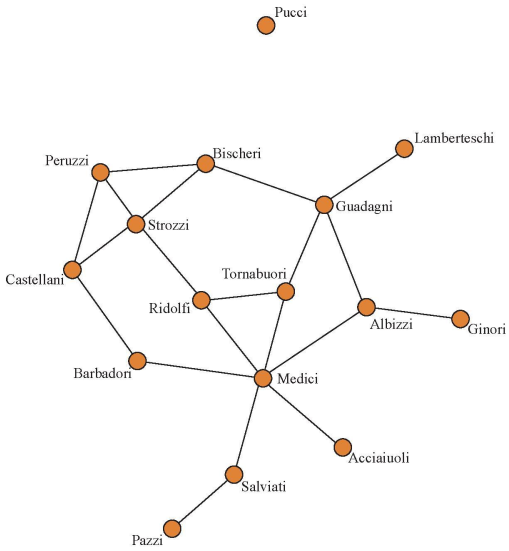

Figure 1: Network of 15th century Florentine marriages by Padgett and Ansell (1993)

establishing frequent communication and trust between two families back in the 15th century, the Medici family turns out to have been the main connector of non-adjacent families, i.e. families that were not directly connected by marriage. In fact, they facilitated the connection between more than 50% of all pairwise family combinations in the network. According to this concept, which is known as betweenness centrality and formally introduced in section 2.3, the other families relied on the Medici family to spread information or reach political consensus which explains the influence of the family. This example illustrates how network analysis can assist in breaking down the interrelations between many different actors of a particular system.

A sports team can also be understood as a complex social network consisting of a set of agents interacting with each other (Ramos et al., 2018). That is why SNA was introduced to the analysis of team sports. It is able to model and characterize their interaction structure and the impact and influence of individual players to the overall team performance (Duch et al., 2010). Players are formally represented by nodes and the interactions or communication between them by edges or ties.

Existing studies have followed different approaches to model the interactions between players. McLean et al. (2019) focus on the direct verbal communication between players, while Sasaki et al. (2017) model the joint defensive actions to represent interactivity. However, successfully played passes are the most popular approach to model the interactive behavior between members of the same team, thus far (Ribeiro et al., 2017; Sarmento et al., 2018). As a common performance indicator in team sports, passes appear to be a natural choice to model team interactions (Hughes & Bartlett, 2002).

Nodes, on the other hand, either represent the actual players by name, playing positions or zones on the pitch, in which passes are controlled or executed by a player. Codifying player names suggests a practical match analysis to evaluate the contribution of actual players to the interplay of a team as executed by various studies (e.g. Cotta et al., 2013). In comparison, tracking playing positions or zones increases the possibility to compare network metrics across multiple matches as done by Clemente and Martins (2017a).

Figure 2 shows two exemplary graphs that model the passing interactions of two opposing football teams across a single match (Pena & Touchette, 2012). Currently, the two most frequently used sources for the interaction data are aggregated passing matrices, which are publicly available, alongside a more time consuming video analysis of matches (e.g. McLean et al., 2018). In football, aggregated interaction matrices, reporting the passes between the players of both teams throughout an entire match, are provided by the UEFA Champions League and FIFA World Cup. They set the basis for the network analysis of matches in several studies (e.g. Clemente & Martins, 2017b).

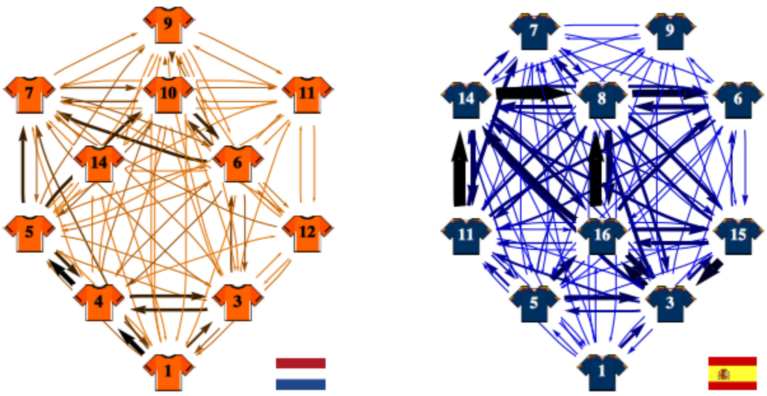

Figure 2: Illustration of two passing networks as modeled by Pena and Touchette (2012)

Based on the idea of passing networks, existing studies have focused on the identification of the individual contribution of players in the interplay and analysis of the overall passing structure of teams to characterize performance. As an example, Gama et al. (2014) investigate the involvement of individual players in football matches and discover that only a subset of them is responsible for the majority of passing. Grund (2012) assesses the general passing structure of football teams and detects a positive link between a balanced distribution of passes and successful performance outcomes.

Overall, SNA has proven its potential to model the complex interactive behavior of sports teams in order to facilitate a better understanding of the nature of team sports and to provide relevant feedback to coaches.

1.3 Critique of Current Research

However, existing studies on SNA in team sports show certain limitations and reveal research gaps that this dissertation aims to address. This section outlines the critique of the current research which is structured into three parts.

1. SNA, as a method to describe team performance, is almost exclusively applied to football.
2. Passing networks are predominantly applied at match-level limiting the informational value of the analysis.
3. Network approaches hardly consider the dynamics of passing interactions.

1. Limited application to other team sports

First, SNA, as a method to evaluate team performance, is almost exclusively applied to football. Only a limited set of studies explore the interaction patterns and most involved players in other team sports such as basketball (Fewell et al., 2012; Clemente et al., 2015a). This might be partly attributed to the worldwide popularity and the low-scoring nature of football. The latter aspect encourages researchers to find alternative performance indicators because scoring opportunities might not be considered as reliable performance measures (Passos et al., 2016). Yet, other team sports like basketball, handball or hockey are similarly complex due to their invasive nature and number of players that need to coordinate their actions (Lemmink & Frencken, 2013). In these invasion games, teams aim at invading the territory of the opponent team through ball interactions in order to score. At the same time, they are trying to prevent the opposing team from scoring (Gréhaigne & Godbout, 1995). Thus, performance heavily depends on the interactive coordination between team members and the application of SNA can contribute to the understanding of interplay in these team sports. Moreover, each team sport faces different sets of constraints. According to Araujo and Davids (2016), this leads to different interaction patterns being needed in order to succeed. Hence, SNA can also uncover how differences in constraints lead to variations in the interplay structures across different team sports. This potential of the method has not yet been exploited.

2. Predominant application at match-level

Second, SNA is almost exclusively conducted at match-level, thus far. This implies an analysis based on one snapshot of the interaction network that reflects all passing interactions of a team aggregated across every ball possession or play throughout a match.[2] In fact, the majority of studies quantify network metrics based on the aggregated passing data of a match captured in an adjacency matrix. This might be partly induced by the easy availability of aggregated interaction data at match-level, e.g. by the FIFA or UEFA in football. Moreover, the computation of popular network metrics such as closeness centrality require each player to be included in the passing interaction which is more likely when looking at an entire match.[3] This leads to several limitations that need to be addressed.

On the one hand, there is no breakdown to individual attacking plays that shape the goal-oriented interactions under varying situational contexts. The match-level approach does not provide an understanding of interplay in different contexts such as variations in the starting location of a play, varying tactical formations or constraints such as the relative location of the goal, players and the ball. However, these factors influence the decision-making on passing and, thus, the overall coordination patterns of a team (Araujo et al., 2006). Thus, SNA applied to teams sports needs to adjust its level of analysis in order to provide an understanding of the emerging interaction patterns in a variety of situational contexts. Only then, coaches can evaluate and classify the performance of their teams and adjust training drills accordingly (Gómez et al., 2013).

On the other hand, the match-level approach prevents a clear and rigorous identification of the relevant interplay that leads to successful or unsuccessful performance. Many existing studies focus on the performance process of teams which is often referred to as the style of play (Grund, 2012; Clemente, 2018). However, there is also an interest in what characterizes successful and unsuccessful performance, following a differential approach. Eventually, this may provide coaches with insights on the interaction patterns and contributions of players that lead to success (McGarry, 2009). Pina et al. (2017) aggregate passes from successful and unsuccessful plays across 15min-time intervals, while Grund (2012) and Clemente (2018) connect match-level metrics with match outcomes such as total shots on goal, the number of scored goals and winning, drawing or losing the match. Although these studies follow a differential approach, they cannot isolate the interplay of specific attacks that is relevant for the outcome of

[2]Henceforth, a play refers to a single ball possession of a team consisting of at least one successful pass.

[3]Again, a more formal introduction to specific network metrics is provided in chapter 2.

a match, especially in low-scoring team sports like football. According to Ramos et al. (2018), the analysis needs to be broken down to separate plays, also referred to as a play-level approach, and the respective interplay connected to the corresponding play-level outcomes which drive the overall performance in team sports. There are some studies that focus on a play-level approach already. However, they do not differentiate between successful and unsuccessful performance (e.g. McLean et al., 2017a). By only looking at successful plays, e.g. those that lead to scoring a goal, there is just a limited possibility to assess the variation in individual metrics across playing positions. This is due to the absence of a comparison value per playing position measurable from an analysis of plays that do not lead to success. This would enhance the suitability of the applied network metrics as valid performance indicators (Sampaio & Leite, 2013). Moreover, these studies only focus on the overall team structure and neglect the contribution of individual players. Hence, a play-level study that regards successful as well as unsuccessful performance is required, with a special focus on the contribution of individual players.

3. Limited consideration of dynamics

Third, irrespective of a match- or play-level perspective, current approaches analyze static snapshots of networks that disregard the dynamics of interactions in team sports. In particular, the sequential order of passes during ball possessions cannot be reconstructed from the static perspective that graphs and matrices provide. However, in order to adequately describe team performance, the dynamic nature of interactions and, thus, team performance, needs to be considered (Sampaio & Leite, 2013). Currently applied network metrics are not only unable to capture the temporal order of passing sequences, some of the frequently used individual metrics, e.g. closeness and betweenness centrality, even make implicit assumptions concerning the dynamics on the passing network (Ribeiro et al., 2017). However, these do not necessarily reflect the sequential order of passes because they are derived from an optimization process at match-level. In particular, the metrics are built on the assumption of network flow as introduced by Atkin (1974) and outlined in detail in section 2.2. According to Atkin, flows are what actually passes between nodes based on a defined network structure. As an example, understanding the connections and relationships between a group of people is often just the basis to assess how information could be spread through the social network. In the analysis of the influential families of Florence in the 15th century, the importance of the Medici family was assessed by its strategic position to facilitate the flow of information between other families. To be precise, the calculated betweenness metric counted the frequency with which the Medici family was able to bridge the shortest path between other families in order to pave the flow of information, i.e. via the shortest number of mediators.

In a sports context, players with a high betweenness value are currently attributed the quality of a bridging player that connects players of its team in ball possession phases (e.g. Clemente et al., 2016a). However, these studies are based on static networks that simply aggregate the passing interactions across plays. More specifically, the shortest path between two players at match-level does not necessarily reflect a ball possession in its sequential order of passes. Figure 3 illustrates how the application of betweenness centrality can lead to inaccurate sport-specific deductions. The figure shows an exemplary graph representing the aggregated passing network between three players. It is based on two separate ball possession phases as shown on the far left in their sequential order of play. In its current application, betweenness centrality would identify bridging players based on the aggregated network. According to this procedure and the underlying directed network, each player would be assigned the same bridging quality as they equally often bridge the flow or connection between the other two players. However, this conclusion turns out to be inaccurate when considering the actual order of passes in each play. In fact, player A does not connect the other two players in any play, being positioned either at the start or end of a play. Moreover, while player C is acting as an actual bridging player in both plays, player B fulfills that role only in one of them. Therefore, there is no justification for equal betweenness values from a performance perspective in sports. Thus, there can only be a metaphorical or approximate assignment of the bridging player role when using the traditional betweenness metric. Likewise, the application of closeness centrality leads to the same fallacy.

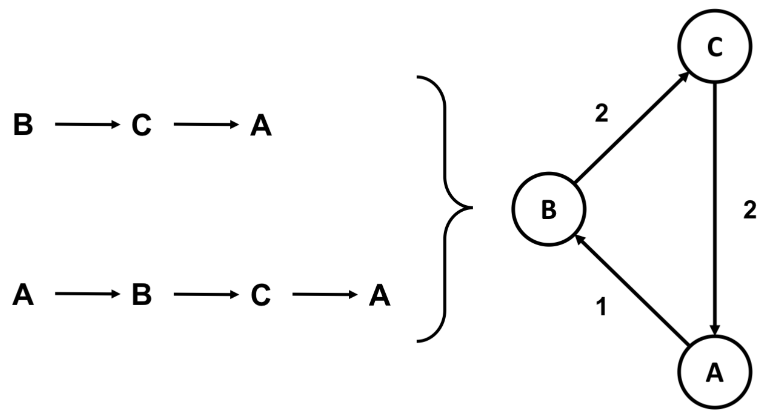

Figure 3: Exemplary passing network to illustrate flaws in current betweenness application

In order to bridge the gap between SNA and performance analysis, the dynamic nature of interactions and, thus, team performance, needs to be taken into account when applying network metrics (Sampaio & Leite, 2013). In particular, as many studies apply the traditional metric of betweenness centrality, there is a need for an alternative metric that assesses bridging players while regarding the sequential order of passes.

1.4 Aims of Dissertation

This section outlines the aims of this dissertation built on the critique of the current research that applies SNA to evaluate performance in team sports. Overall, the goal is to extend and improve the existing work of SNA applied to passing interactions in team sports. From a theoretical point of view, this aims at a more adequate modeling of performance considering situational contexts, performance outcomes and the dynamics of interactions. From a practical or applied perspective, the more accurate quantification of team performance can potentially elevate the insights generated from practical match analysis and, thus, lead to a better performance feedback for coaches, analysts and players. The specific aims are subdivided into three pillars which are outlined in the following. They are also connected to the corresponding studies that define this dissertation.

1. The application and extension of SNA to other team sports than football.
2. The practical implementation of play-level analysis to model interplay in different situational contexts and with different performance outcomes.
3. The consideration of the dynamic nature of interactions in team sports.

1. Extension to more team sports

The first aim of this dissertation is to extend the analysis to more invasion team sports beyond football. Thus far, nearly all studies apply SNA to football. However, invasion team sports, in general, are described as complex dynamic systems consisting of interacting parts, as outlined previously (Davids et al., 2013). Hence, the goal is to improve the understanding of the interactive behavior of players in additional team sports. A second aim is to also contribute to the general understanding of invasion team sports. By offering a comparison of the interaction patterns in the different sports, the effect of varying constraints on interplay, which are induced by the different natures of each invasion team sport, can be assessed. Among these are differences in the rules of the game, the number of allowed players in a team or pitch sizes and markings.

2. Practical implementation of play-level analysis

Apart from an extension to other team sports, this dissertation also aims at the practical implementation of a play-level analysis. This is because match-level approaches can neither differentiate between interplay in varying situational contexts nor provide a clear and rigorous identification of the relevant interplay that leads to successful or unsuccessful performance. Thus, this dissertation aims at the implementation of practical studies that breakdown the analysis to individual plays in order to i) achieve a tactical analysis that compares interplay in different situational contexts and ii) statistically evaluate the connection between interplay patterns and successful as well as unsuccessful performance outcomes.

3. Consideration of interaction dynamics

Third, this dissertation aims at the consideration of the dynamic nature of interactions in team sports. Thus far, network metrics have been transferred, almost unchanged, to the analysis of interaction networks in team sports. The frequent application of betweenness centrality, in particular, currently leads to a violation of the actual dynamics in team sports. To initiate the consideration of the dynamic nature of interactions, the aim is to develop a new metric that assesses bridging (or intermediary) players by analyzing actual passing sequences and, thus, regard the sequential order of passes.

Connection to studies of dissertation

The outlined aims of the dissertation were pursued in three studies. Table 1 visualizes the classification process to underline the contribution of each study to the realization of each aim of this dissertation. The first study aimed at the application of SNA to more team sports beyond football. For the first time, SNA is applied to handball and a team-level analysis conducted in basketball. The study included a comparison of their respective interaction patterns to obtain a better understanding of invasion team sports as such. The second study aimed at the practical implementation of a play-level analysis in handball. It specifically targeted the analysis of interplay in varying situational contexts through a consideration of changes in the attack formations. As SNA had not been applied to handball prior to this dissertation, the study also built on the first aim of this dissertation. The third study aimed at a play-level analysis in football applying metrics that consider the sequential order of passes while offering a connection to successful as well as unsuccessful performance outcomes. Thus, it contributed to the second part of the practical implementation of a play-level analysis as well as the consideration of interaction dynamics, pursuing the final aim of this dissertation.

Aims of dissertation	Study 1	Study 2	Study 3
1. Extension to more team sports	✔	✔	
2. Implementation of play-level analysis			
Consideration of situational context		✔	
Connection to performance outcomes			✔
3. Consideration of interaction dynamics			✔

Table 1: Classification of studies in line with aims of dissertation

1.5 Outline of Dissertation

The remainder of this dissertation is structured as follows. Chapter 3 provides a deeper insight into SNA and its application in team sports. This is followed by a presentation of the three studies that define this dissertation in chapter 4. It includes a short overview on each study alongside its embedding into the literature highlighting the contribution and novelty. Chapter 5 discusses the theoretical and practical impact of the dissertation on performance analysis in team sports. Moreover, the chapter highlights still existing limitations of the network approach that are present in this dissertation. Chapter 6 provides an outlook and concludes.

2 Methods

This chapter provides an exhaustive overview on SNA including its application and interpretation in a team sports context. Thus, it is meant to deepen the knowledge on the method acquired during the introduction. Moreover, it targets a better understanding of current network modeling and analysis of performance in team sports as well as the network procedures used in the three studies of this dissertation. Section 2.1 provides a short historical overview on the development of SNA, its diffusion into various fields of research and general application in sports science. This is followed by a more formal introduction of the method in section 2.2. Based on the acquired theoretical understanding, 2.3 presents the most popular network metrics that are commonly applied to passing networks. Both sections combine formal aspects with a team sport-related interpretation to provide a better understanding of the method and its applicability to performance analysis. Finally, section 2.4 provides an exhaustive discussion on the existing studies that apply SNA in team sports.

2.1 Historical Background

The foundation of SNA dates back to the 1930s and is attributed to Jacob Moreno and Helen Jennings who were interested in the link between individual well-being and interpersonal relations. Hence, they modeled social relationships between individuals as a network and visualized them in a graph. This approach, which the two researchers labeled as psychological geography, became later known as sociometry (Scott, 1988). The first ever published article in the field studied the interpersonal relations between school girls in New York City using sociograms as illustrated in Figure 4 (Moreno, 1934). The aim of the study was to visualize the different structures of interpersonal relations and, building on that, deduce a connection to the mental state of school children, e.g. how social isolation is connected to depression or the feeling of loneliness.

In the 1940s, a research group at Harvard University, led by William Warner and Elton Mayo, built on the idea of sociograms to explore interpersonal relations in work environments. The aim of the research group was to observe actual group behavior at work, independent of the formal organization, in order to study the efficiency of working teams (Roethlisberger & Dickson, 1939). Their research pioneered the assessment of team performance applying a network model.

Starting in the 1950s, the until then qualitative procedures were complemented by a more mathematical approach to formalize the method. Bavelas (1950) introduced

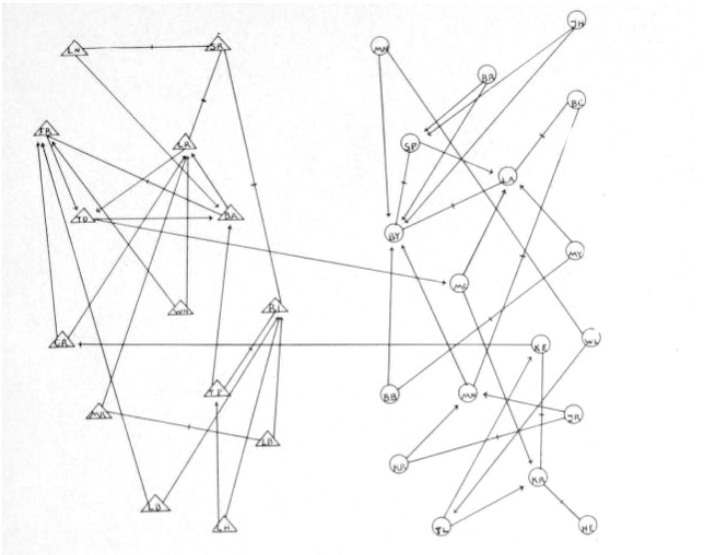

Figure 4: Sociogram of friendships in a NYC elementary school by Moreno (1934)

the idea of structural centrality to quantify the strategic position of individual nodes and the resulting influence on the overall network. In addition, Austrian psychologist Siegfried Nadel was the first researcher who brought forward a mathematical approach to quantify the global properties of network structures (Nadel, 1957).

In the 1970s, Granovetter's work on the importance of weak ties fostered the bridging process between qualitative sociograms and the quantitative analysis of interrelations. In his work, he outlined how novel information and ideas can only be found outside of closely related groups or cliques (Granovetter, 1977). Moreover, Freeman extended the previous work of Bavelas by formalizing and introducing a set of centrality measures that quantify the role and position of nodes in a network (Freeman, 1978).

Induced by the formalization process and theoretical developments of SNA, the 1990s saw an increasing transfer of the method to other research fields beyond sociology. Considering the research fields with the highest applications of the method, Newman proposed a categorization of different network types. He distinguished between social networks that focus on concepts such as friendship or collaboration, information networks that model diffusion of information, technological networks that are rather of physical nature such as logistics networks, and biological networks that model nano-level processes such as the interaction between proteins in the human body or neural networks (Newman, 2018).

Based on the work by Roethlisberger and Dickson (1939), there has also been a growing interest in using SNA to analyze the communication and organization of work teams in order to evaluate the ideal organizational setup to jointly perform tasks.

Bavelas (1950) examined ideal structures for group performance and found that complex tasks were best executed in a centralized setup with a single leader coordinating actions, while decentralized structures performed better on simple tasks. Balkundi and Harrison (2006) detected that diversified communication between work colleagues lead to increased team performance. In their analysis of email traffic in virtual teams, Gloor et al. (2006) attributed good performance to a balanced communication between colleagues.

The evaluation of team performance using SNA naturally paved the way of the method into sports science. While there is a body of research in sports management science focusing on the network structures in sport organizations such as the collaboration between franchises or the communication within sports associations (Cousens & Slack, 1996; Seevers et al., 2010; Ratten et al., 2011), network studies in sports science address actual team performance in sports. Here, one can differentiate between two approaches, following Wäsche et al. (2017). Competition networks apply SNA to establish a performance ranking through the comparison of competition outcomes between players or teams. Lai et al. (2018) model match outcomes between professional Italian table players in a network, as visualized in Figure 5, to identify the best players in the country. By considering indirect connections, their aim was to compare the strength of players that had not actually played against each other, similar to the Elo rating system for relative performance evaluation in chess. In tennis, Radicchi (2011) took a similar approach to detect the best tennis players of all-time. Dey et al. (2017) compared the performance of cricket players to establish a performance ranking and optimal team composition.

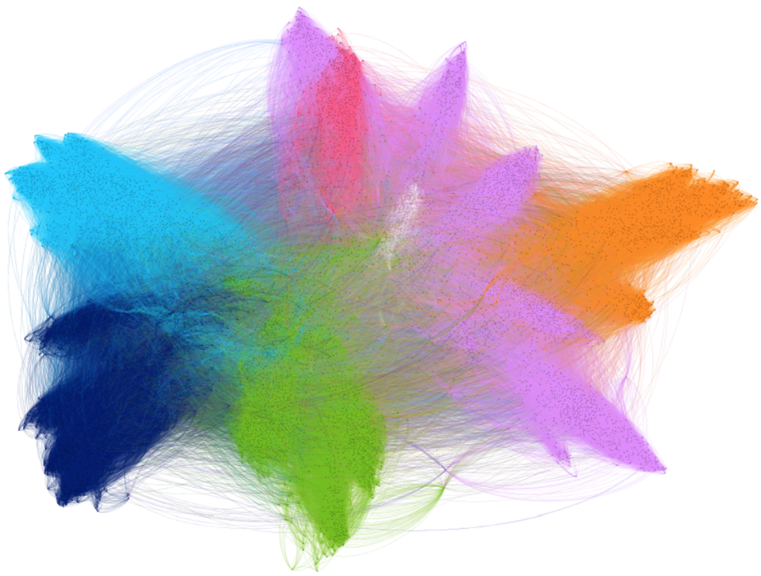

Figure 5: Network visualization of table tennis matches played in Italy by Lai et al. (2018)

In contrast, interaction networks focus on the performance of sports teams by modeling the actual interactions taking place between players of a team during a match (Wäsche et al., 2017). Different types of interactions are modeled in existing studies. In the majority of studies, passes are used to model the interaction between players during a match (Duch et al., 2010; Passos et al., 2011; Grund, 2012). Other less frequently used interaction types are the verbal communication within a team during a match (McLean et al., 2019), the joint defensive actions of team members as analyzed by Sasaki et al. (2017) in rugby, and the positional changes taking place between players (Passos et al., 2011).[4]

2.2 Network Modeling

After an illustrative introduction to SNA, including a brief historical background, this section takes a rather formal approach by introducing the basic concepts and terminology of the method. This is accompanied by a sport-related interpretation to provide a more targeted understanding of the ideas and concepts. Moreover, the concept of network flow, which was briefly discussed in the introduction, is revisited and explained in detail. Flows model the dynamics on a network and many network metrics applied to team sports are built on the very concept. Thus, a rigorous understanding assists in evaluating the implicit assumptions of network metrics against the actual dynamics on passing networks in team sports.

Basic concepts and terminology

A network consists of a set of nodes connected by ties that link them (Borgatti & Halgin, 2011). More formally, following Wasserman and Faust (1994), a set of n nodes, referred to as N, can be denoted by $N = \{n_1, n_2, ..., n_n\}$. Then, x_{ij} defines a tie between nodes n_i and n_j, where $i \neq j$, often referred to as edge or link. In the most general case, x_{ij} can take on two states:

$$x_{ij} = \begin{cases} 1, & if\ there\ exists\ a\ tie\ between\ nodes\ n_i\ and\ n_j \\ 0, & otherwise \end{cases} \tag{1}$$

If we revisit the Florentine marriage example from the introduction, x_{ij} indicates whether any set of two families had been connected by marriage or not. In a team

[4]For the purpose of this dissertation, the remainder of this chapter will focus on interaction networks that model passes between players.

sports context, a tie indicates whether a successful pass was played between players n_i and n_j or not.

Building on this basic binary model of connections, network models are often extended to a directed and/or weighted version. Ties between any set of two nodes might not necessarily be reciprocal. In an exemplary setting, person A might like person B but the affection is not reciprocated. Therefore, directed networks differentiate between a tie going from node n_i to n_j and the opposite direction. Then, x_{ij} can take on a different value than x_{ji}. This provides a deeper insight into the underlying relationship between two nodes. In a team sports context, modeling a directed network allows a differentiation between the direction of a pass, revealing which player is passing the ball and who is on the receiving end.

Weighted networks consider the intensity of a tie or frequency of an interaction. While marriage or friendship are rather binary concepts, the number of meet-ups between two friends is not. In that case, x_{ij} is not restricted to 0 and 1 anymore. To be more rigorous, one can differentiate between state-type ties and event-type ties in weighted networks. Weighted state-type ties describe connections with a certain continuity and intensity to them (Borgatti & Kidwell, 2011). Examples are commonly found in logistics networks, e.g. when modeling the distance between production sites or distribution centers as ties. There is a natural steadiness to the distance between locations. In contrast, weighted event-type ties have a discrete nature that can be counted over periods of time (Borgatti & Halgin, 2011). Then, accumulating the frequency of interactions between nodes defines the strength of the underlying tie. Examples for event-type ties are mails sent between co-workers or the number of meet-ups between friends.

In a team sports context, weighted networks allow the modeling of the frequency of passes between any set of two players going beyond the pure representation of the existence of a connection. Thus, passes are weighted event-type ties as they are countable over periods of time defining the strength of the connection between players which, then, define the passing network structure of a team. The accumulation process of the interactions can be conducted over different time frames such as one passing sequence. In football, for example, this can be defined from the first pass of a ball possession until the ball is either lost, the game interrupted or a goal scored (Pollard & Reep, 1997). This procedure represents the play-level approach as introduced in the first chapter. Similarly, passing interactions can be accumulated over certain time intervals or the entire match. The accumulation of interactions between players can also be done according to a certain filter throughout a match such as counting all passes from ball possessions that started in a particular zone on the pitch. Each approach yields different network structures.

Figure 6 summarizes the four most common forms of network modeling in exemplary graphs. In line with graph theory, the dots represent nodes, while the lines connecting the dots model the ties between them. Thicker lines represent stronger connections, while arrows indicate the direction of a connection or relationship. The bottom left network visualizes an undirected and unweighted network. In a sports setting, this form of modeling is able to represent which players of a team interacted with each other ignoring any details on the passing direction or intensity. The upper left graph represents a directed but unweighted network. Building on the previous example, it can model the direction of the passes, though with no information on the frequency of the event. The bottom right network models an undirected but weighted network which can reflect the passing intensity between any set of two players, though not the direction of the passes. The upper right network depicts a directed and weighted network completing the set of options. Following the previous example, this model can capture the direction and frequency of passes between players.

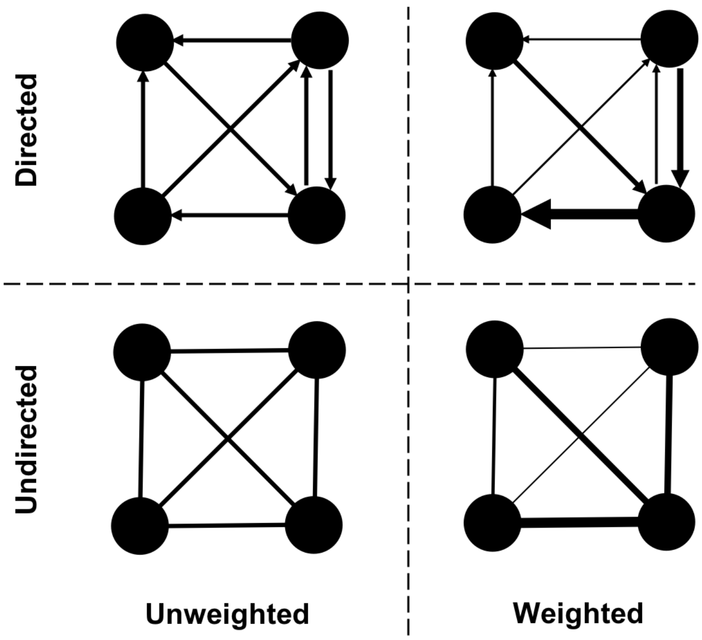

Figure 6: Exemplary directed and weighted networks

Apart from visualizing networks in graphs, representing a rather qualitative approach, the information on connections between nodes is commonly captured in a $n \times n$ adjacency matrix representing the ties between each pair of nodes. In such a matrix the element of the i-th row and j-th column is given by x_{ij} and any form of directed or weighted network model can be represented. The diagonal elements of the matrix are usually zeros assuming that no node is connected to itself. This assumption also makes sense when analyzing passing networks in team sports. Figure 7 shows an exemplary graph and its corresponding adjacency matrix.

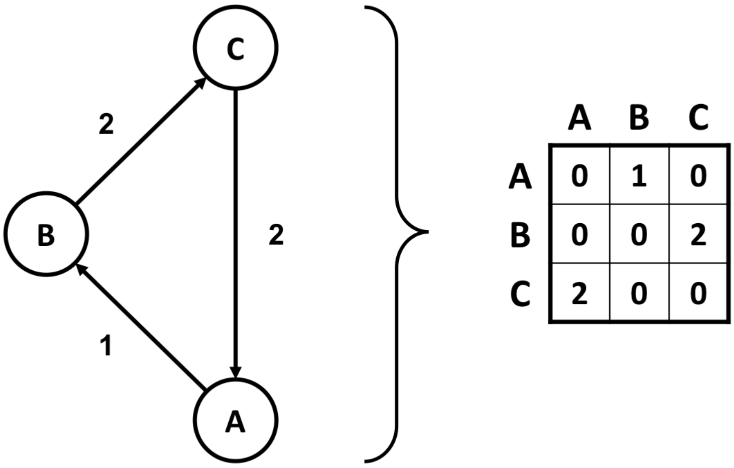

Figure 7: Exemplary network representation in matrix form

Dynamics on networks

The interactions between nodes are modeled as ties that define the network structure. However, in SNA, ties are often only seen as the basis for studying the dynamics on the network. Atkin (1974) outlined this idea by describing ties as the backcloth or pipes to facilitate flow across the network structure. Ties are seen as channels to transfer resources, either of material or non-material kind (Wasserman and Faust, 1994). Thus, flows represent the dynamics on the network.[5] While marriages symbolized the official ties between families in the Florentine marriage example, Padgett and Ansell were interested in the resulting dynamics on the network structure, i.e. the flow of information between families. As briefly outlined in the introduction, many network metrics that are applied in the analysis of team sports are built on the idea of flow. In the following, the implicit assumptions they make concerning the dynamics on the networks they are applied to are evaluated against the dynamics on passing networks in team sports.

Technically, there are two common types of flow, walks and path. They describe different forms of trajectories that a flow follows along edges and nodes on a particular network structure (Borgatti, 2005). In fact, they determine the rules of the flow through the network (Borgatti & Halgin, 2011). Walks impose no rules on how resources are transferred from one node to another through the network structure.

[5]For clarification purposes, these are different to the dynamics *of* the network which study changes in the network structure itself (Ramos et al., 2018).

This means that nodes and edges can be revisited multiple times, e.g. a person might give a banknote to a store and receive it back the next day due to another purchase (Borgatti & Halgin, 2011). In contrast, paths connect nodes without revisiting any nodes or edges twice. In particular, geodesics are flows that resemble the shortest path between two nodes on an existing network structure.

Figure 8 visualizes the different types of flows in an unweighted network setting. For illustration purposes, the respective trajectories of flow are marked in color. While the graph on the far left resembles an exemplary walk that revisits node B, the exemplary path between nodes A and D in the middle graph does not revisit any node or edge. The graph on the far right, depicts the shortest path or geodesic between nodes A and D which, obviously, is facilitated by the direct tie between the two nodes. This demonstrates why it does not make sense to model the flow of information as a walk in the Florentine marriage example. Intelligence is only useful if supplied for the first time and is unlikely to be passed on twice via the same link. Thus, Padgett and Ansell applied network metrics that model the flow of information as geodesics.

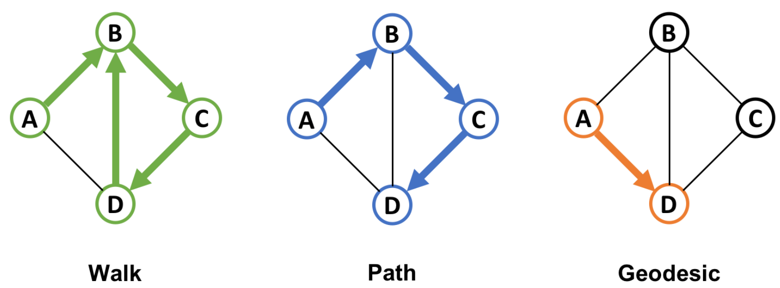

Figure 8: Exemplary walk, path and geodesic

As mentioned before, marriage is of course a binary concept leading to an unweighted network structure. In a weighted network model, e.g. a traffic network, shortest paths usually reveal the quickest or cheapest route between cities across an existing network of roads. Here, the sum of edge weights, e.g. reflecting the total travel duration in hours, is minimized on the shortest path. In that case and opposite to the example in Figure 8, the direct road connection between two cities might not necessarily resemble the shortest path and, thus, route of flow. As an illustrative example, let the graph in Figure 9 resemble a traffic network consisting of cities A, B and C which are connected by road. Here, edge weights denote the respective travel durations between them in hours. Then, the shortest path between cities A and C, and thus travel duration, is in fact facilitated via city B. This is because the sum of the two edge weights is smaller (5 hours in total) than the edge weight of the direct

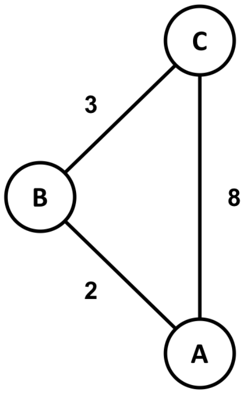

Figure 9: Example of a shortest path in a weighted network

connection between cities A and C (8 hours in total). Here, a traffic jam might have caused the delay on the direct road connection between the two cities. The important take away is that in a weighted network model, routes of flow such as geodesics are determined by edge weights.

In a sports context, the concept of shortest paths or geodesics between players must be interpreted differently. Strong connections, represented by the amount of passes between players, are rather seen as favorable in sports. In contrast, traditional application areas value small edge weights, e.g. reflecting shorter travel durations. Thus, the reciprocal values of edges are commonly considered to adjust for the minimization problem and, hence, detect shortest paths in a passing network (Newman, 2001). Figure 10 visualizes the idea in an exemplary passing network between players A, B and C. Now, there have been two successful passes between players A and C. Thus, the edge weight between them is given by the reciprocal value ($\frac{1}{2}$ or 0.5). Then, as there were ten passes between players A and B and 20 passes between players B and C, the summed reciprocal value of the two ties that connect players A and C, indirectly, is $\frac{1}{10} + \frac{1}{20} = 0.15$. This is smaller than the reciprocal value obtained from the direct connection. Thus, according to Newman's procedure, the shortest path between players A and C is facilitated via player B. In current network studies on team sports, this is how shortest paths are determined.

However, shortest paths do not reflect the actual sequential order of passes because they are determined based on the aggregated passing network across all plays throughout a match. In fact, any route of flow based on accumulated passing data does not necessarily reflect the interplay as it unfolded. Therefore, the adaption of the flow concept is not as straightforward in a sports setting. As outlined previously, some commonly applied network metrics, e.g. betweenness centrality, are built on the assumption of flow. That way, they assign dynamics on the network that do not reflect

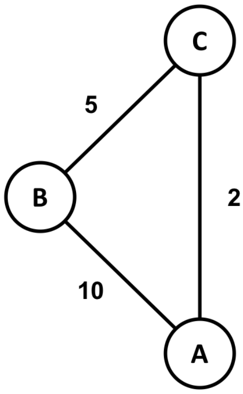

Figure 10: Exemplary of shortest path in a passing network

the sequential order of passes and, hence, disregard the actual performance process in team sports. The next section will discuss these and other network metrics in detail.

2.3 Network Metrics

Having introduced the concepts and terminology of network modeling, this section provides an overview on network metrics. They offer a quantification of the properties of a network based on the interaction data captured in an adjacency matrix. As such, they complement the visualization of networks in graphs. On a micro- or individual-level, network metrics focus on the position of individual nodes within the network by assessing their centrality or clustering tendencies. In contrast, macro- or team-level metrics focus on the assessment of the global structure of the underlying network.

In a sports context, network metrics aim a quantification of the performance process in team sports. Micro-level metrics quantify the individual contribution or role of a player in the overall interplay. At a macro-level, network metrics analyze the passing structure of a team as a whole. The position and role of players as well as the passing structure is statically assessed based on one snapshot, most commonly reflecting the passing interactions of a team throughout a match (Ramos et al., 2018).

This section outlines the most prevalent metrics in network studies on team sports. It provides a formal introduction as well as an interpretation in a sports context. The descriptions are occasionally accentuated with exemplary graphs for illustration purposes.

2.3.1 Individual-level metrics

On an individual-level, network metrics assess the position of a node within an existing network. This allows a comparison between nodes and quantification of their contribution to the overall network structure (Jackson, 2010). In a team sports context, individual-level metrics assess the role and importance of individual players in the interplay of the entire team.

Technically, this is carried out by evaluating the centrality or clustering tendencies of a node. The evaluation of centrality is commonly subdivided into four different metrics (Borgatti & Everett, 2006): i) degree centrality measures how connected a node is directly with other nodes, ii) eigenvector centrality assesses the importance based on how central the immediate neighborhood of a node is, iii) closeness centrality assesses how strong the connection is to all other nodes in the network - directly and indirectly, and iv) betweenness centrality captures how important a node is based on how well it facilitates the indirect connections between the other nodes in a network. Betweenness and closeness centrality are determined based on a node's role in the network flow on an existing network structure as discussed preciously. In particular, both metrics apply the concept of shortest paths to assess the importance of nodes.

Apart from centrality, a different approach to assess a node's position is to evaluate its clustering tendencies. This implies an assessment of the level of connectedness in the node's immediate neighborhood. The underlying metric is called clustering coefficient and is outlined in detail in the following alongside the centrality measures that were introduced above.

Degree centrality

Degree centrality assesses the direct connections of a node by counting the number of edges between itself and the other nodes of the network. Thus, as the simplest measure, it assesses the centrality of a node by its immediate connections as defined by Proctor and Loomis (1951) and Shaw (1954). The degree centrality index, $C_D(n_i)$, for node n_i in an unweighted and undirected network is calculated as,

$$C_D(n_i) = \sum_{i \neq j} x_{ij} \tag{2}$$

where x_{ij} defines the tie between nodes n_i and n_j. In a directed network, in-degree counts the incoming edges of a node, while out-degree aggregates the outgoing edges. This differentiation allows a more accurate evaluation of the connections. In a weighted network, the intensity of the direct connections is also considered resulting in the weighted degree metric.

In a team sports context, degree centrality determines the importance or centrality of a player by assessing the direct interactions with team members. If modeled as an unweighted and undirected network, the focus lies on the simple count of different team members that a player passes with at least once. In a directed but unweighted network setting, there is a differentiation between the number of team members that a player successfully passed to and received the ball from at least once. If the passing interaction is modeled as a weighted network, the frequency of passes is considered as well.

Figure 11 is an exemplary graph visualizing the snapshot of a passing network between five football players. The degree centrality of player A is, $C_D(A) = 4$, as he interacts with all four team members at least once. His in-degree centrality is, $C_{ID}(A) = 3$, because player A receives at least one pass from three players (B, C and E) and the out-degree centrality is, $C_{OD}(A) = 4$, because he passes the ball to all four players at least once. The weighted in-degree value of player A aggregates all successfully received passes ($C_{WID}(A) = 15$), while the weighted out-degree value counts all successfully played passes by player A ($C_{WOD}(A) = 10$).

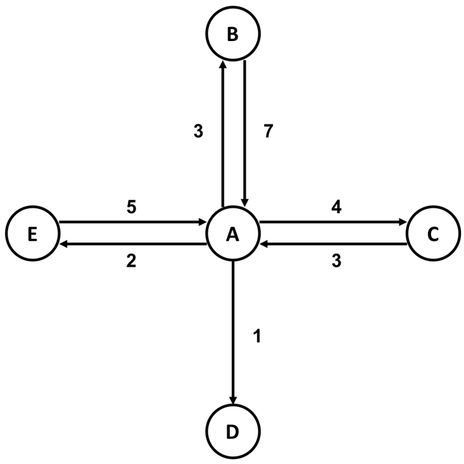

Figure 11: Exemplary passing network for degree measurement

For comparison purposes across different passing networks, the metric can be normalized by the maximum number of team members that a player can potentially interact with (unweighted network) or by the player's share in the overall passing intensity of the team (weighted network). In summary, degree centrality addresses the most direct form of involvement in the interactions of a team and does not rely on any flow processes on a given network structure.

Eigenvector centrality

Eigenvector centrality evaluates the importance of a node based on how central all adjacent nodes are. Similar to degree centrality, it measures the centrality of a node based on the number of ties it has with other nodes in the network. However, as an extension, the importance of the node is not only dependent on the number of direct connections but also on the connectivity of the adjacent nodes, as proposed by Bonacich (1972). Thus, the metric assumes that not all connections are equally important following the belief that influential adjacent nodes increase one's own influence (Newman, 2016). The eigenvector centrality index, $C_{EV}(n_i)$, for node n_i in an unweighted and undirected network is calculated as,

$$C_{EV}(n_i) = \frac{1}{\lambda} \sum_{i \neq j} x_{ij} \; C_{EV}(n_j) \tag{3}$$

where x_{ij} defines the tie between nodes n_i and n_j, $C_{EV}(n_j)$ is the eigenvector value of node n_j and λ is a constant. Thus, centrality is measured as a function of the centrality of all adjacent nodes. By construction, the metric cannot be applied to weighted and directed network models.

In a team sports context, the metric attributes high centrality to players that interact with many team members who themselves have many direct passing connections. Thus, passing solely with one particular player who is highly connected within the rest of the team might lead to larger eigenvector centrality than interacting directly with many unpopular players, in terms of their inclusion in the passing.

As the interaction can only be modeled in an undirected and unweighted setting the intensity and direction of passing is ignored in the estimation. In order to obtain any meaningful variation in the eigenvector centrality values across players in a team, the network must not be complete. A complete network would imply that all players are connected with each other leading to equal eigenvector centrality values by definition.

Closeness centrality

Closeness is a measure of centrality first introduced by Bavelas (1950) to assess how well-connected a node is with all other nodes in a network. The concept is based on the belief that a node is central if it is influential within its network. This is evaluated by its proximity to all other nodes (Wasserman & Faust, 1994). Technically, this is calculated by the inverse sum of the shortest distances between the node of interest and all other nodes in the network. The closeness centrality index, $C_C(n_i)$, for node n_i in an unweighted and undirected network is calculated as,

$$C_C(n_i) = \left[\sum_{i \neq j} d(n_i n_j) \right]^{-1} \quad (4)$$

where $d(n_i n_j)$ denotes the shortest distance between nodes n_i and n_j. In an unweighted setting, the shortest distance between two nodes is given by the total number of edges to be traversed on the shortest path between them. In a weighted network, the distance is calculated by the sum of all edge weights. A technical requirement and limitation of closeness centrality is that all nodes need to be reachable within the network.

In a team sports context, the immediate translation is how well a player is interconnected with all other team members or how much he influences the overall passing network. The evaluation is based on a snapshot of all played passes during a certain amount of time and strong connections are seen as favorable as outlined in the previous section. Then, the player who minimizes the distance to all other players, measured by the inverse sum of reciprocal passing values, has the highest closeness centrality and is, thus, regarded as most central.

Figure 12 is an exemplary graph visualizing the snapshot of a passing network between four football players. The edge weights represent the number of passes played to each other. As outlined in the previous section, strong connections are seen as favorable in a sports context. Thus, the reciprocal values of edge weights should be considered. Then, player D has the highest closeness centrality. The minimum distance between him and player A is $\frac{1}{5} = 0.2$, player B is $\frac{1}{4} + \frac{1}{10} = 0.35$ and player C is $\frac{1}{4} = 0.25$. The corresponding centrality value is $C_C(D) = \frac{1}{0.35+0.25+0.2} = 1.25$. In contrast, player A has the lowest closeness centrality ($C_C(A) = \frac{1}{0.55+0.45+0.2} = 0.8\overline{3}$).

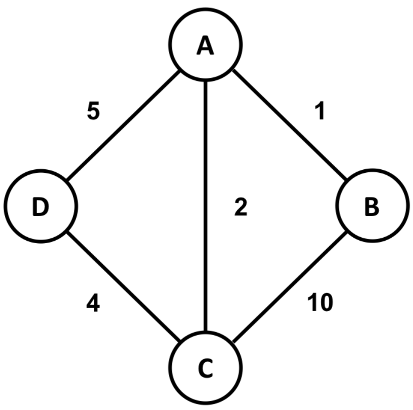

Figure 12: Exemplary passing network for closeness measurement

As closeness centrality is built on the concept of flow, shortest paths between players do not necessarily resemble actual passing sequences. Therefore, there is rather a metaphorical than sport-specific interpretation to closeness centrality.

Betweenness centrality

Betweenness centrality is a measure first introduced by Freeman (1977) that assesses the importance of a node by its intermediary role in bridging the connection between nodes of the same network. More specifically, the metric counts how often a node lies on the shortest paths between any set of two nodes, evaluating centrality by its bridging quality. Thus, the node with the best strategic position to facilitate flow between nodes is regard as most central. Betweenness centrality index, $C_B(n_i)$, of node n_i in an unweighted and undirected network is calculated as,

$$C_B(n_i) = \sum_{j \neq k \neq i} \frac{g_{jk}(n_i)}{g_{jk}} \tag{5}$$

where g_{jk} denotes all shortest paths between nodes n_j and n_k, while $g_{jk}(n_i)$ represents a subset of those paths in which node n_j is functioning as a bridging unit. The distance between nodes is evaluated either by the number of edges or the sum of edge weights.

In a team sports context, the immediate translation would be how often a player is bridging the strongest pairwise passing connection between members of his team. Again, Figure 12 can serve as an example to illustrate the idea of betweenness centrality in a sports context. Closeness centrality was interested in the total distances of the shortest paths between players. Betweenness centrality assesses how often a player is on the shortest paths between any set of two team members. In the given example, this is the case in two occasions for players C and D, respectively. The shortest path between players A and C is via player D, the shortest path between players B and D is via player C, and the shortest path between players A and B is via both, players C and D. Following Freeman (1977), betweenness values are often normalized by the maximum number of pairs ($\frac{(n-1)(n-2)}{2}$) that the player of interest could potentially be in-between of given a passing network of n players. Thus, the normalized betweenness centrality of players C and D in the given example is $\frac{2}{3}$ equally. Players A and B are not on the shortest paths between any pair of players resulting in a betweenness score of 0.

Again, the shortest path between any set of two players does not necessarily reflect actual consecutive passes of the ball. Hence, players with high betweenness centrality do not necessarily connect other players during a ball possession. The calculation is based on the aggregated passing data captured during a certain amount of time. Thus, betweenness centrality rather provides a metaphorical than sport-specific interpretation of an intermediary or bridging player.

Local clustering coefficient

The local clustering coefficient is a measure that indicates how closely interconnected the direct neighborhood of a node is. Following Watts and Strogatz (1998) and Fagiolo (2007), it calculates the proportion of adjacent nodes that are also directly connected with each other. That way, one can evaluate whether the node of interest is part of a subgroup that frequently interacts with each other leading to strong triangular connections. Local clustering coefficient index, $CC(n_i)$, of node n_i in an unweighted and undirected network is calculated as,

$$CC(n_i) = \frac{\sum_{j \neq k \neq i} x_{ij}\ x_{ik}\ x_{jk}}{C_D(n_i)\ (C_D(n_i) - 1)} \tag{6}$$

where x_{ij} defines the tie between nodes n_i and n_j and $C_D(n_i)$ the degree centrality of node n_i. In a weighted and directed setting, the intensity and direction of the connections are also considered. In contrast to the previously introduced centrality metrics, the local clustering coefficient evaluates the immediate structural properties around the node of interest.

In a team sports context, it is a measure to evaluate whether a player is part of a subgroup that frequently interacts with each other. Similar to betweenness centrality, this does not necessarily imply that triangular relations are formed sequentially. Calculating clustering coefficients is a static evaluation based on an existing network structure of played passes that considers the overall interactions between players.

2.3.2 Team-level metrics

Team-level metrics assess the structure and properties of a network as a whole. In a team sports context, they assess the general structure of interplay between the entire team. Total links, reciprocity and density assess the general connectivity between all nodes in the network. Centralization and heterogeneity assess the differences between the individual-level metrics of all nodes in the network. Moreover, the average value of a particular micro-level metric across all nodes of a network can also be considered as an indicator for the general network structure. However, this straightforward procedure is not further investigated in this section.

Total links

The total links of a network offer an understanding of the general intensity of interactions between nodes in a network. As the simplest measure, they are the summation of all edges between nodes in a network. Total links index, TL, is calculated as

$$TL = \sum_i \sum_j x_{ij} \tag{7}$$

where x_{ij} defines the tie between nodes n_i and n_j. The idea can also be extended to a weighted and directed network setting. In an unweighted setting the values are bounded between 0 and the number of possible connections ($\frac{n\ (n-1)}{2}$) given a network with n nodes. In a weighted setting there is no upper boundary.

In a team sports context, this metric signals the total number of passing connections between players (unweighted network) or the sum of all passes played (weighted network).

Reciprocity

Reciprocity evaluates to what extent the connections between nodes are mutual in a network. Thus, it signals how mutually balanced or one-sided the connections of a network are. It is measured by the proportion of reciprocal connections against the total number of links. Reciprocity index, R, is calculated as

$$R = \frac{\overleftrightarrow{TL}}{TL} \tag{8}$$

where $\overleftrightarrow{TL}$ signals reciprocal connections, while TL denotes the total number of connections as introduced before. Thus, values are bounded between 0 and 1. The lower bound indicates that there are no reciprocal connections at all, while the upper bound implies that all connections are reciprocal. In a weighted setting, reciprocity evaluates to what extent the intensity of a connection is reciprocated.

In a team sports context, reciprocity measures to what extent interplay and the intensity of interplay between players is mutual within a team. In case the overall passing network is considered, the metric does not reflect actual interplay in form of one-two passes, i.e. situations in which player A passes to player B and back to player A. Reciprocity indicates how balanced the overall passing between players A and B is across a certain amount of time.

Density

Density measures the general level of cohesion in a network. Following Wasserman and Faust (1994), the metric evaluates the number of actual connections against the number of possible connections between nodes in a network. Density index, D, is calculated as,

$$D = \frac{2\ TL}{n\ (n-1)} \tag{9}$$

where TL is the number of actual connections and $\frac{n\ (n-1)}{2}$ the number of possible connections. Thus, the metric is bounded between 0 and 1.

In a team sports context, density quantifies the level of cohesion and, thus, traces the degree to which pairwise player connections are exploited. Figure 13 provides the example of two passing networks involving four players each. Network a) exploits half of its six potential connections resulting in a density value of 0.5. In contrast, the number of actual connections equals the number of potential connections in network b) leading to a density value of 1. This implies that there is a passing connection between every set of two players.

While density measures the general level of cohesion in terms of passing, there is no information on the concentration of interplay around certain focal players.

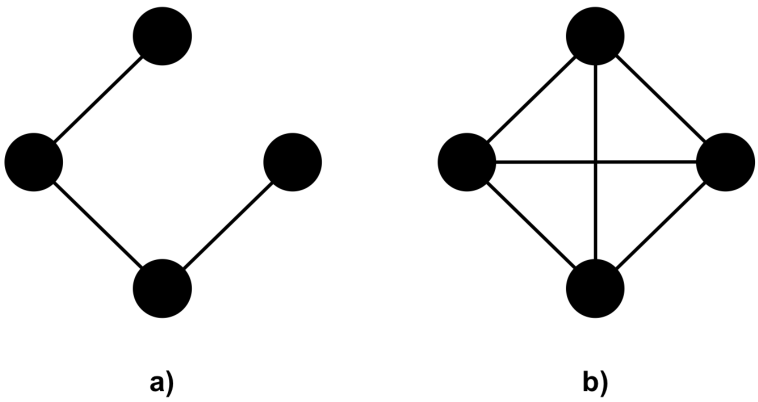

Figure 13: Exemplary passing networks for density measurement

Centralization

Centralization evaluates to which extent the cohesion is concentrated around a focal point or node in a network (Freeman, 1978). Therefore, it can be seen as a complementary measure to density. It measures the normalized deviation of a certain individual-level metric between all nodes and the node with the highest metric value in the network. Using degree centrality as an exemplary metric, the degree centralization index, DC, is calculated as,

$$DC = \frac{\sum C_D* - C_D(n_i)}{(n-2)\ (n-1)} \tag{10}$$

where C_D* is the highest degree value of a node in the network and $C_D(n_i)$ the degree value of node n_i. The summed differences are normalized by the maximum deviation possible ($(n-2)\ (n-1)$), bounding the maximum degree centralization value to 1.

In a team sports context, the metric measures to what extent interplay is dependent on or centralized around one player. The exact type of dependency is determined by the choice of the individual-level metric employed. In the degree case, a high centralization value indicates that the interplay is mostly structured around one focal player, while a centralization towards 0 signals a balanced involvement of all players across a team.

Figure 14 provides the example of two passing networks involving five players each. While both networks have the same density value, $D = \frac{2\times 4}{5(5-1)} = 0.4$, the interplay in network b) is more concentrated around one focal player than in network a). In fact, the degree centralization of network b) is 1, while the value for network a) is only $\frac{1}{6}$.

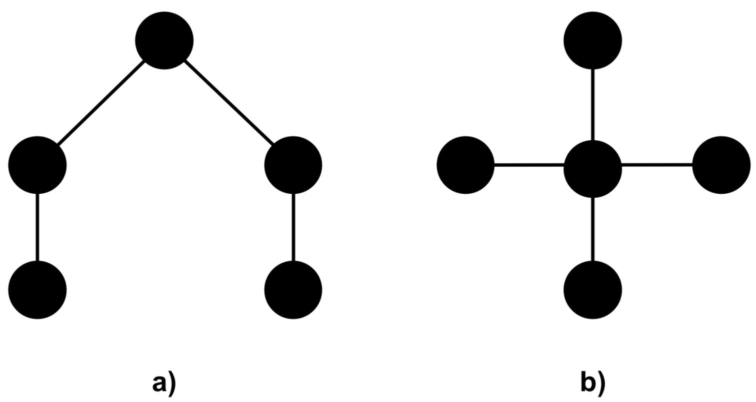

Figure 14: Exemplary passing networks for centralization measurement

The metric assumes that there is only one focal player and is, thus, designed to detect the extent to which the interplay is concentrated around this player. Centralization does not consider the distribution of the passing shares of the other players.

Heterogeneity

Heterogeneity evaluates to which extent the cohesion is concentrated around multiple hubs (Dong & Horvarth, 2007). In contrast to centralization, the metric does not focus on a single focal node but evaluates overall deviation tendencies in a network. Following Snijders (1981), the variance of a certain individual-level metric across all nodes of a network is measured adjusted by the mean value. Using again degree centrality as an exemplary metric, the degree heterogeneity index, DH, is calculated as,

$$DH = \frac{n \sum C_D(n_i)^2 - (\sum C_D(n_i))^2}{(\sum C_D(n_i))^2} \tag{11}$$

where n is the number of nodes in the network and $C_D(n_i)$ the degree value of node n_i. The metric has a lower bound of 0 signaling no prevalence of hubs.

In a team sports context, heterogeneity measures to what extent interplay is dependent on or centralized around certain hubs. Hence, it identifies whether ball possessions are concentrated around a subset of players. In the case of degree centrality, heterogeneity values are increasing with an imbalance in passing shares across players.

Figure 15 provides the example of two passing networks involving six players each. While both networks have the same density value, $D = \frac{2 \times 5}{6(6-1)} = 0.\overline{3}$, the interplay in network b) is more concentrated around a hub, consisting of two players, than in network a). In fact, the degree heterogeneity of network b) is 0.32, while the value for network a) is only 0.08.

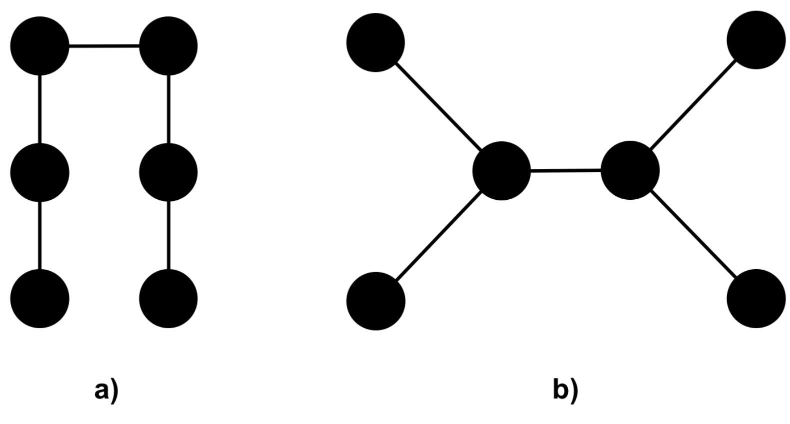

Figure 15: Exemplary passing networks for heterogeneity measurement

While the metric is sensitive to deviations as such, it is not designed to detect any specific number of focal players.

2.4 Related Work

This section provides an extensive overview on the research conducted on passing networks in team sports using SNA. In particular, the section discusses the range of team sports and competitions that SNA has been applied to, the procedures taken to model passing networks and the variety of network metrics and visualization techniques applied, thus far.

Teams sports and competitions

The majority of studies that model interaction networks in team sports have been conducted in football. Only a few studies focus on interaction networks in other team sports. Fewell et al. (2012) and Clemente et al. (2015a) analyze the passing interactions between team members in basketball for multiple matches in the National Basketball Association (NBA) and in a training environment, respectively. Sasaki et al. (2017) model the defensive behavior of rugby players defining the strength of a tie between two players by the number of their joint defensive actions during a match. In other team sports, there is no immediate application of network analysis on the interaction patterns during training or matches, thus far. Thus, the remainder of this section focuses mostly on football as nearly all studies are conducted in this sport.

Here, most samples consist of matches from male professional leagues and major professional tournaments. Clemente et al. (2016a) compare the passing interactions between teams of the Spanish LaLiga and the English Premier League. Grund (2012) focuses on a passing analysis of Premier League matches from multiple seasons, while the study by Gama et al. (2015) provides an example of SNA conducted on matches from the Portuguese Football League.

On a tournament-level, several studies focus on passing networks from matches in the UEFA Champions League (Trequattrini et al., 2015; Clemente & Martins, 2017a; Clemente & Martins, 2017b; Pina et al., 2017; Oliveira & Clemente, 2018). McLean et al. (2017a) analyze the passing interactions of teams participating in the COPA American football championships and European football championships in 2016 setting the focus on national teams. Moreover, there are several studies analyzing passing structures of national teams that participated in FIFA World Cup tournaments between 2010 and 2018 (Pena & Touchette, 2012; Cotta et al., 2013; Clemente et al., 2015b; Clemente et al., 2016b; Peixoto et al., 2017; Clemente, 2018; Praça et al., 2019). Besides, there is also some research on SNA in a controlled football training environment (Gonçalves et al., 2017; Praça et al., 2017; Praça et al., 2018).

Modeling passing networks

The two most frequently used sources for passing data are aggregated passing matrices that are publicly available alongside the data extraction from self-conducted video analysis. Aggregated interaction matrices reporting the passes between players are provided by the UEFA Champions League and FIFA World Cup in football and set the basis for the network analysis in several studies (e.g. Clemente & Martins, 2017b). In contrast, McLean et al. (2018) collect passing data through a video analysis of recorded matches at the European football championships in 2016.

The level of analysis is often determined by the data source. Passing data that is publicly available provides a single interaction matrix per team representing the aggregated passing interactions throughout a match. Thus, SNA is necessarily conducted at match-level. In contrast, video analysis provides a flexible choice as passing sequences from individual ball possessions can be captured in separate matrices. Although the majority of studies are still analyzed at match-level, some studies break down the analysis to intervals (e.g. Pina et al., 2017) or even plays (e.g. McLean et al., 2017a).

Independent of the level of analysis, nodes either represent actual players, playing positions or zones on the pitch in which passes are controlled or executed by a player. Actual player names are mostly used in studies that aim at a practical match analysis to evaluate the role of players in the interplay of their team (Duch et al., 2010, Pena & Touchette, 2012; Cotta et al., 2013; Trequattrini et al., 2015). In contrast, the majority of studies conduct an ex-post codification of playing positions. This procedure increases the comparability of network metrics across multiple matches aiming at a rather general understanding of interplay in the sport. Some studies neither focus on the individual analysis of players nor playing positions. For example, McLean et al. (2018) divide the football pitch into zones to represent nodes.

Application of network metrics

Based on the network modeling, network metrics are applied to quantify the involvement of players and the general interplay structure. Here, the majority of studies apply network metrics at match-level.

Focusing on individual metrics at match-level, weighted in- and out-degree are frequently used as an alternative to reporting overall passing statistics. Weighted in-degree is employed to identify how often the ball is directly passed to a player during a match. The midfield position is identified as the most targeted playing position in the majority of studies (Clemente et al., 2015b; Clemente et al., 2015c; Clemente & Martins, 2017a). In contrast, weighted out-degree is used to identify the playing position that is frequently passing the ball. Similar to weighted in-degree, midfielders and especially central midfielders are attributed the highest centrality based on that particular metric (Clemente et al., 2015 JPES; Clemente et al., 2015b; Clemente & Martins, 2017a; Peixoto et al., 2017). There is also a study by Aquino et al. (2018) that applies eigenvector centrality as an extension to degree centrality. This study also highlights midfielders as most central. Other studies calculate clustering coefficients to assess potential cliques of players that frequently interact with each other. The clustering tendencies turn out to be evenly distributed across playing positions (Clemente et al., 2014a; Gama et al., 2015; Arriaza-Ardiles et al., 2018).

However, the two most applied individual metrics at match-level are betweenness and closeness centrality. Betweenness centrality is used by researchers to detect intermediary players in the interaction process. Players with high betweenness centrality are referred to as bridging players that ensure the ball flow between players (Arriaza-Ardiles et al., 2018). Pena and Touchette (2012) interpret betweenness centrality as the dependency on a player to facilitate ball flow. Clemente et al. (2016c) describe right defenders and forwards as bridging players as they obtain the highest betweenness values in their study. In fact, in the majority of studies midfielders and especially central midfielders are identified as the playing position with the highest betweenness centrality (Clemente et al., 2015b; Clemente & Martins, 2017a; Peixoto et al., 2017; Aquino et al., 2018; Castellano & Echeazarra, 2019).

Closeness centrality is used to identify how well connected a player is within the team. While Clemente et al. (2016c) cannot detect any tendency towards a playing position in their study, Pena and Touchette (2012) ascribe defensive and central midfielders the highest closeness value. This is in line with Clemente et al. (2015b), Aquino et al. (2018) and Castellano and Echeazarra (2019) who identify central midfielders as the playing position with the highest closeness centrality. As the metric requires the inclusion of all players in the network, it is not surprising that it is exclusively applied at match-level maximizing the likelihood of all players being involved in the interplay at least once.

Turning to team metrics at match-level, several studies report the average sum of the weighted in- and out-degree value of each player (Clemente et al., 2015b; Clemente et al., 2016b; Peixoto et al., 2017). The last-mentioned study finds this metric to be significantly negatively correlated with the number of scored goals. The less passes are played and received by each player on average, the more goals are scored by that particular team. Clemente (2018) calculates the fraction of links that is reciprocal. As he models an aggregated passing network, the weighted reciprocity coefficient is used.

Moreover, density and total links are computed in the majority of studies to better assess the general level of cohesion in football matches. While Peixoto et al. (2017) find the metric to be negatively correlated with the number of scored goals by a team, Clemente (2018) detects a significantly positive correlation of the density value of a team with its total number of shots on goal during a match. In addition to the general level of cohesion, several studies focus on the extent to which the cohesion is concentrated around focal players or hubs. Clemente et al. (2015d) and Gama et al. (2015) apply weighted degree centralization and degree heterogeneity to detect focal players. Their results suggest a rather balanced passing involvement across all players in football. Grund (2012) calculates weighted in- and out-degree centralization to assess whether the number of passes played and received is concentrated around a focal player, respectively. He finds that centralization is negatively correlated with the total number of goals scored by a team. This suggests that the absence of a focal player is favorable in terms of team performance.

The discussed studies were all conducted at match-level, calculating individual and team metrics based on the interactions of a team across an entire match. A minority of studies assesses smaller interaction networks. Interval-level analysis investigates the passing interactions across certain time-intervals of a match, while play-level analysis looks at each ball possession separately. On an interval-level, Clemente et al. (2016a) capture six passing networks per team per match representing 15-min intervals. They identify the midfield positions to be most involved in the interplay measured by weighted in- and out-degree. The same result is obtained by Yamamoto (2010) who analyzes one football match using 5min-intervals. Turning to team metrics, Pina et al. (2017) calculate density, the average clustering coefficient and degree centralization for twelve Champions League matches using 15min-intervals. Following the idea of using the entering of a finishing zone as a proxy for the successful outcome of a ball possession (Tenga et al., 2010), they find density to be a significant predictor for successful play outcomes. According to their results, reduced density corresponds to a larger amount of successful plays.

On a play-level, McLean et al. (2017a) apply density and reciprocity to analyze ball possessions that lead to scoring a goal. McLean et al. (2018) also calculate in- and out-degree in a separate study. However, they do not model players or playing positions as nodes, but define four equally large pitch zones in which the ball was either controlled or passed from. Moreover, the concept of flow centrality is used to assess the involvement of players in the interplay at play-level. Instead of analyzing the mean involvement of a player in each ball possession across a match, flow centrality builds the ratio of plays that a player is part of to the total number of plays of its team. Thus, each ball possession phase is looked at separately and the metric is bounded between 0 and 1 by construction. Using this metric, Clemente et al. (2014b) identify the right defender position to be most involved in football matches. Duch et al. (2010) define flow centrality as the fraction of plays in which a playing position is involved in that lead to a shot on goal. They identify midfielders to be most involved. In contrast, Fewell et al. (2012) calculate flow centrality in basketball. They identify the point guard position as most central in terms of its overall involvement across a match.

Visualization of passing networks

Evidently, SNA in team sports is primarily facilitated by the calculation of individual and team metrics. However, it is often augmented by the visualization of the underlying passing networks as illustrated in Figure 2 of the introduction. In most studies, the aggregated interactions between players across a match are visualized and the arrangements of the nodes resemble the tactical formations of teams. The direction of the connections is displayed using arrowed lines, while the intensity of the connections is expressed by the thickness of the lines. Duch et al. (2010) extend the representation of passing networks by also modeling possession outcomes as can be seen in Figure 16. A directed connection towards a possession outcome either represents a shot on goal or off target. In the play-level study by McLean et al. (2017a), the sequential order of the passes of one ball possession phase is visualized. In contrast to the other representations that visualize aggregated networks, their focus lies on visualizing the passing sequences as they unfold on the pitch.

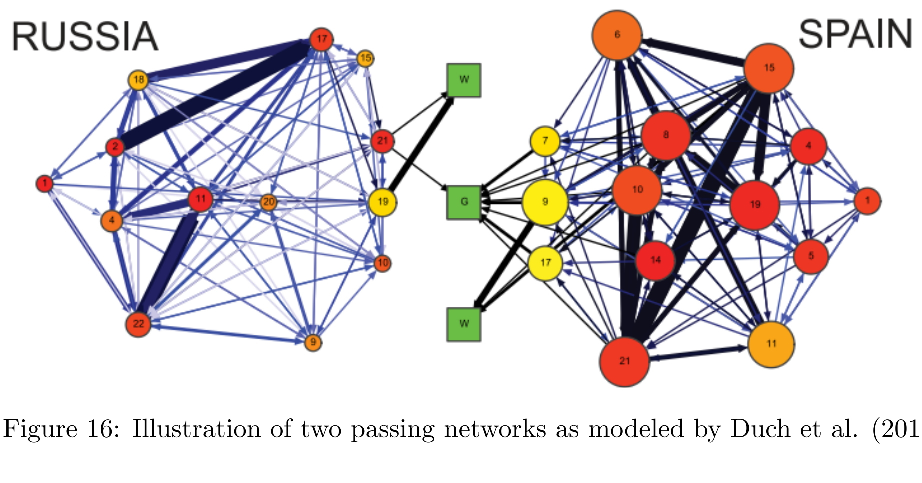

Figure 16: Illustration of two passing networks as modeled by Duch et al. (2010)

3 Studies

This chapter outlines the studies defining this dissertation, consisting of three journal articles (two published and one in press) and their specific contribution based on the outlined aims in the introduction. They all focus on interaction networks using passes to model ties. Moreover, nodes represent playing positions which is in line with the majority of research conducted in the field. The first study addresses the scarcity of team sports that network analysis has been applied to alongside a comparison of interaction patterns across sports. The second study is a practical implementation of a play-level analysis in handball to enable a comparison of interplay in different situational contexts. The third study focuses on football and addresses the limited consideration of interaction dynamics while connecting network measures to performance outcomes at play-level. The original and full version of each article can be found in the Appendix.

3.1 Study 1 - Characterizing different team sports using network analysis

Citation

Korte, F., & Lames, M. (2018). Characterizing different team sports using network analysis. Current Issues in Sport Science, 3:005.

Contribution of author

The author of this dissertation was the principal investigator and author of the accepted article. He developed the idea for the study, the study design, and chose the methods to be used. Moreover, he led the statistical procedures and interpretation of the data. Martin Lames supported the development of the research design and the statistical procedures. The author of this dissertation wrote the article, accepted in Current Issues in Sport Science, while receiving feedback from his co-author.

Summary

The main motivation for the first article was to extend the application of network analysis to more team sports by modeling their passing interactions as a network. Here, the aim was not to create an in-depth performance analysis of each examined sport. Instead, the goal was to generate first insights on the interaction patterns of various invasion team sports and comparing the basic structure of interplay and dominant playing positions between the sports. As team sports differ in the constraints that players face, different patterns of interaction are required in order to succeed (Araújo & Davids, 2016). Thus, this study did not only aim at a better understanding of the interaction patterns within each sport, but also at uncovering and quantifying the resulting differences between them.

Eight matches from professional men's tournaments in football, basketball and handball were examined, respectively. For simplicity reasons, a match-level approach was chosen analyzing the overall passing interactions of each match. As the focus of the study lied on the variety of team sports and comparison of interplay between them, the most common individual metrics of the existing literature were applied, namely weighted in- and out-degree, closeness and betweenness. The aim was to assess the dominant playing positions of each team. The codification of playing positions enabled the comparison between centrality values across matches. Different types of centralization metrics were also employed to identify to what extent interplay could be concentrated around a focal player.

Moreover, minimum spanning trees (MSTs), previously applied in finance to uncover asset correlations in the stock market (Bonanno et al., 2003), were introduced to the interaction analysis in team sports as a visualization technique. Visualizations of passing networks in previous studies appeared rather overloaded by the abundance of connections that are accumulated during a match. MSTs only visualize the strongest passing connections under the side condition that each player is still part of the resulting interaction network.

The defensive midfielder in football, the point guard in basketball and center in handball were identified as most the prominent players in the interplay of their sport. For football and basketball, this was in line with previous literature (Fewell et al., 2012; Clemente & Martins, 2017a). As it was the first application of network analysis in handball, there were no comparison values. The tendencies towards a focal player were highest in handball, followed by basketball and football. The application of MSTs uncovered the basic structure of interplay in the respective team sports. While in football the strongest interplay was visible between defensive players and along attacking wings, a star topology centered around the point guard became visible in basketball. In handball, the MST resembled the basic tactical formation in attacking plays suggesting that individual centrality values might be partly predefined by the playing position in handball.

Overall, this was the first study that drew a comparison of the passing interactions and dependencies on dominant playing positions between different team sports. For the first time, a network approach was taken for the analysis of passing structures in handball and team properties in basketball. Moreover, the study contributed to the understanding of the similarities and differences between invasion team sports using their interplay structures. In line with previous studies, it highlighted how different constraints, such as the rules of the game or markings of the pitch, affect the interaction patterns of invasion team sports. The introduction of MSTs facilitated an alternative and comprehensible visualization of interplay structures at match-level.

3.2 Study 2 - Passing network analysis of positional attack formations in handball

Citation

Korte, F., & Lames, M. (in press). Passing network analysis of positional attack formations in handball. Journal of Human Kinetics.

Contribution of author

The author of this dissertation was the principal investigator and author of the accepted article. He developed the idea for the study, the study design, and chose the methods to be used. Moreover, he led the statistical procedures and interpretation of the data. Martin Lames supported the development of the research design and the statistical procedures. The author of this dissertation wrote the article, accepted in Journal of Human Kinetics, while receiving feedback from his co-author.

Summary

The main motivation for the second article was the practical implementation of a play-level analysis in handball. In that particular sport, the passing interaction during positional attacks, defined as the collective offensive actions of a team in which all playing positions are taken, is crucial to achieve a favorable throwing position on goal (Wagner et al., 2014). Building on the findings from the first study, passing networks in handball were further explored. Attacks are frequently played in different formations with variations in the numbers of attacking and defending players. This is induced by temporary suspensions as well as the permission to replace the goalkeeper for an additional offensive player. Thus, the second study specifically targeted a practical play-analysis of interplay to account for varying situational contexts.

22 professional handball matches from the European Men's Handball Championship 2018 were analyzed. In comparison to previous network studies in football,

playing positions were directly tracked instead of tracking players in first place and assigning playing positions ex-post. This was due to the more static setup of players in handball (Korte & Lames, 2018). Moreover, tracking playing positions allowed a fluent observation process due to the frequent substitutions in handball. In addition, Gwet's AC1 inter-rater statistic was added to the reliability testing of the data. Complementary to Cohen's Kappa, Gwet's statistic analyzed the agreement on the actual occurrence of a pass before turning to the passer and receiver in detail. This was due to the rapid nature of interplay in handball prone to the oversight of passes during video analysis.

Weighted in- and out-degree, flow centrality as introduced by Fewell et al. (2012), density and degree centralization were computed to assess the involvement of players and general structure of interplay in the four most prevalent attack formations in handball. Frequently used metrics such as betweenness and closeness centrality were not employed due to their assumptions on the network that do not necessarily reflect actual interplay in handball. The used network measures made no requirements on the size and connectivity of the network.

Regardless of the attack formation, interplay was found to be structured around the center and back positions. This was in line with the results of the first study. However, significant differences were found in the passing contribution of playing positions and interplay structures depending on the attack formations.

Overall, this was the first study conducting an in-depth network analysis in handball at play-level. It provided deeper insights on the passing involvement of playing positions in different attack formations. While the first study dealt with differences in the interaction patterns due to the varying constraints across different team sports, the differences in interaction patterns in the second study were induced by the varying constraints within the sport itself and revealed by a play-level approach.

3.3 Study 3 - Play-by-play network analysis in football

Citation

Korte F., Link D., Groll J., & Lames M. (2019). Play-by-Play Network Analysis in Football. Front. Psychol. 10:1738.

Contribution of author

The author of this dissertation was the principal investigator and author of the published article. He developed the idea for the study, the study design, and chose the methods to be used. Moreover, he led the statistical procedures and interpretation of the data. The data collection was conducted with the support of Johannes Groll. Martin Lames supported the development of the research design and the statistical procedures. The author of this dissertation wrote the article, published in Frontiers in Psychology, while receiving feedback from his co-authors.

Summary

The main motivation for the third article was the execution of a play-level analysis in football considering the sequential order of passes while offering a connection to successful as well as unsuccessful performance outcomes. In particular, the study proposed a metric that detects intermediary players and, thus, offers an alternative to betweenness centrality. In addition, the goal was to connect the resulting network metrics with successful and unsuccessful play-level outcomes. Following Ramos et al. (2018), previous approaches at interval- and match-level did not pin down the relevant interplay that lead to the overall results.

70 professional men's football matches from the 1. and 2. German Bundesliga during the 2017/2018 season were analyzed. As a novel approach and contrary to a video analysis or provision of passing matrices at match-level, action feeds on successfully played passes were combined with positional data on every player and the ball. This facilitated the sequential tracking of passing sequences. Moreover, the differentiation between successful and unsuccessful performance at play-level was approximated by evaluating the entering of a finishing zone during plays, following Tenga et al. (2010) and Pina et al. (2017).

Following Fewell et al. (2012), flow centrality was calculated to assess the fraction of plays across a match that a playing position is involved in at least once. In addition to that, a new metric called flow betweenness was introduced. It measures the fraction of plays across a match that a particular playing position is actually in-between of. In general, that is the case if a player receives and passes on the ball at least once during a ball possession phase, connecting two different team members. Thus, the metric offers an alternative to betweenness centrality at play-level by actually considering the sequential order of passes.

Central defenders were identified as most involved according to flow centrality and most in-between according to flow betweenness. However, this was only the case in unsuccessful plays. Central offensive midfielders were most involved in successful plays, while central defensive midfielders were most intermediary.

Overall, this was the first study that assessed the centrality levels of playing positions in football at play-level while considering the sequential order of passes. The introduction of flow betweenness offered a metric that assesses the intermediary role of players reflecting actual interplay. In general, flow-based metrics were positioned as suitable network measures in football. This is because they are robust to plays consisting of only few passes and players and consider the sequential order of interplay. Moreover, this study offered a differentiation between centrality measures in successful and unsuccessful plays advancing the understanding of the role of playing positions in the interplay of football.

4 Discussion

Building on existing work, the studies of this dissertation extend and refine the modeling of interaction patterns in team sports using SNA. In particular, the studies focused on the extension of the method to more team sports, the breakdown of interplay to individual plays in order to consider varying situational contexts and the modeling of dynamics in passing networks. Thereby, the studies of this dissertation aimed at contributing to the theoretical and practical aspects of performance analysis in team sports.

Figure 17 provides an overview on the studies by classifying them according to their level of analysis and consideration of dynamics. Study 1 was conducted from a static perspective at match-level and focused on the general application of SNA in more invasion team sports beyond football. In fact, it is the first study that applied SNA in handball. Moreover, it pioneered the comparison of interaction patterns, both on individual- and team-level, between different team sports. The complexity of the interaction networks, due the abundance of interactions across an entire match, was reduced using MSTs. Study 2 addressed the limited context that was provided at match-level in the first study by facilitating a breakdown of passing networks to separate plays in handball. The procedure demonstrated how play-level analysis regards the varying situational contexts that impact team interactions and, thus, performance. Study 3 modeled the sequential order of plays in football by introducing flow betweenness to consider the dynamics of team interactions. Besides, the study connected the play-level analysis with varying performance outcomes to rigorously assess the individual contribution of players.

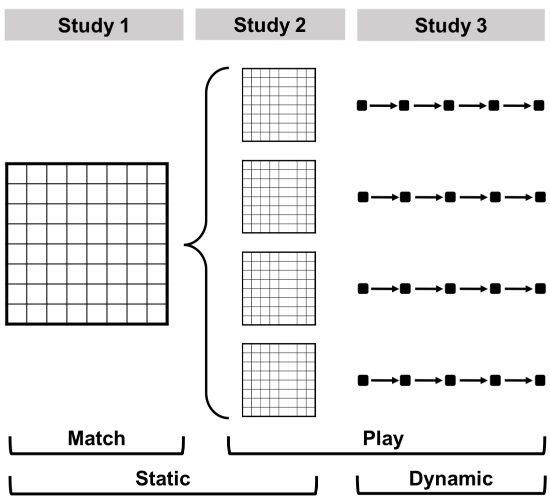

Figure 17: Overview on studies by level of analysis and modeling of dynamics

In the following, the theoretical and practical impact of the studies is highlighted alongside a discussion on the still existing limitations of the network approach in this dissertation.

4.1 Theoretical Impact

Theoretical performance analysis aims at determining the structure and even nature of performance (Lames, 2018). In team sports, this implies a formalized understanding of game behavior including the interactions between players of a team (McGarry, 2009). Thus, the theoretical impact of this dissertation can be evaluated by its contribution to the modeling and structuring of performance in team sports. It is outlined in the following and subdivided into four parts. The first part discusses the overall impact of this dissertation on the performance structuring within and across team sports. The second part focuses on the specific theoretical implications of the play-level approach. The third part evaluates the impact on the modeling of dynamics in the analysis of team performance. The fourth and final part deals with the graph-theoretical impact of this dissertation.

Performance structure within and across team sports

Team sports are characterized by the flow of interactions between players to collectively outperform the opponent team (Mateus, 2005). Interactions patterns emerge under the influence of specific constraints and are crucial in order to succeed (Araujo & Davids, 2016). SNA models the interactive behavior between players that team performance emerges from. Hence, it is self-evident that the mere extension of the method to more team sports beyond football extends the repertoire of modeling performance in the respective sports.

In fact, this dissertation pioneers the modeling and analysis of interplay patterns in handball. This implies a theoretical contribution to the characterization of individual and team performance. On an individual level, most studies in handball have focused on physical profiles, e.g. total running distance in a match, and throwing statistics to characterize the performance of players (e.g. Chaouachi et al., 2009). According to Volossovitch (2013), there is a lack in the modeling of position-specific playing performance to describe the individual contribution of players in handball. Thus, this dissertation adds to the characterization of individual performance in handball by modeling and quantifying the contribution of playing positions to the performance process of a team. On a team level, research has focused on the scoring efficiency of teams to explain performance outcomes (e.g. Vuleta et al., 2003). The collective actions of handball teams, in order to create a favorable throwing position, had not been modeled prior to this dissertation. Thus, the application of team metrics and MSTs also contribute the modeling of collective team performance in handball.

In basketball, the interaction patterns between players of a team had not been studied extensively, either. However, in contrast to handball, there are studies that model the individual involvement of playing positions in successful attacks (e.g. Fewell et al., 2012). Moreover, Bourbousson et al. (2010) assess the degree of coordinated movement of basketball teams. Yet, this dissertation modeled the structure of interplay for the very first time by applying team metrics and MSTs. This facilitates an alternative and complementary approach to modeling team performance in basketball.

In addition to that, this dissertation contributes to the general understanding of invasion team sports by drawing a quantitative comparison between their specific interaction structures and concentration around focal players. The overarching aim of invasion team sports is to collectively outperform the opponent by overcoming defensive lines in order to score. Differences in constraints, such as the rules of the game, players per team or pitch sizes and markings, shape unique interaction patterns across team sports. This dissertation provides a better understanding of invasion team sports as such by focusing on the sport-specific constraints that shape unique interactive behaviors under the sports-covering aim of collectively outperforming the opponent team.

Impact of play-level approach

The second study implemented a play-level approach to model the interplay in varying situational contexts. The third study applied this level of analysis to distinguish between the interaction patterns that lead to successful and unsuccessful performance outcomes. While previous studies filter aggregated interactions across a match by factors such as tactical formations or match status (Clemente & Martins, 2017a; Praça et al., 2019), the breakdown to plays offers a natural subdivision into the smallest units of interplay and, thus, reduces the level of abstraction in the analysis. As the approach can account for different situational contexts, a more precise modeling of team performance is established. As an example, the analysis of handball in different attack formations deepens the understanding of interplay by identifying the relevant factors that structure and influence the performance process. Furthermore, analyzing the passing interactions under the consideration of the novel goalkeeper rule provides a better general understanding of the sport.

Moreover, a play-level analysis yields smaller networks in terms of the number of involved players and amount of interactions in a play. Thus, this dissertation also contributes to the proposition and selection of network metrics that are suitable at play-level. In the study on handball, the majority of individual and team metrics are computed for each play resulting in an average value per match. While the average number of passes ranges between five and seven (depending on the attack formation) in handball plays from the underlying sample, 50% of all plays consist of two passes or less in football (Tenga et al., 2010). Thus, flow-based metrics are identified as suitable for the analysis of the individual contribution in football at play-level because they focus on the proportional prevalence in plays across a match. Overall, the theoretical impact lies in the modeling of interplay at a lower abstraction level and under the consideration of specific contexts as well as the selection of sport-specific network metrics to characterize team performance.

Modeling the dynamics of interactions

Previous research focused on the usefulness of SNA to model the complex interactions in team sports. However, the dynamic nature of interactions has barely been addressed (Ramos et al., 2018). The dynamic perspective on SNA traces the sequential pattern of interactions and exact ball flow between players to adequately model performance in team sports. While the play-level approach is the basis for a dynamic analysis, it still builds on the aggregation of passes and does not guarantee the consideration of the sequential order. With regard to dynamics, the theoretical impact of this dissertation lies in the proposition of flow betweenness which actively considers the sequential pattern to trace actual bridging players. By adjusting to the nature of ties in team sports, it offers an alternative to the traditional betweenness metric.

It should be noted, that all metrics used in study 2 and 3 do not violate the dynamic nature of interplay. Although degree and flow centrality (individual-level) as well as density and centralization (team-level) take a static perspective on networks, they are not built on the traditional assumption of flow. Instead, they model walks of length 1 which reflects a single pass. This does not qualify them as metrics that can trace the sequential pattern of passes, but, in contrast to closeness or betweenness centrality, they do not ignore the actual dynamic nature at play-level. Therefore, a secondary contribution, in terms of dynamics, lies in the selection of network metrics at play-level that, although not actively modeling passing sequences, regard the dynamic nature of interplay. Metrics that rely on the flow assumption, such as betweenness centrality, are classified as approximations of performance in this dissertation that deliver no direct sport-specific interpretation.

Visual representation of interaction networks

Apart from the computation of a novel metric, this dissertation also contributes to the graph-theoretical aspects of SNA applied to team sports. Thus far, the visualization of passing networks had mainly focused on the depiction of the overall interactions between players at match-level (e.g. Pena & Touchette, 2012). As performance analysis aims at the reduction of complexity and intuitive representation of performance (Passos et al., 2016), the introduction of MSTs to highlight the strongest connections and basic structure of interplay extends the repertoire of visualization tools in performance analysis.

4.2 Practical Impact

The practical impact of this dissertation must be evaluated by its potential usefulness and support for coaches and analysts of sport teams (Carling et al., 2005; Lames & McGarry, 2007). They rely on objective information from past performances in order to plan and conduct practice (Maslovat & Franks, 2008). This requires a profound understanding of the nature of the underlying sport (McGarry, 2009). Moreover, an adequate interpretation and classification of the observed performance is crucial for the practical considerations of training as performance is subject to varying contexts and situational aspects (Lames & Hansen, 2001). This dissertation contributes to the improvement of both aspects.

The mere introduction of SNA to invasion sports such as handball already leads to a better information basis for coaches on the performance of their respective team. That is because the sport heavily relies on passing to achieve a favorable throwing position and, in a practical match analysis, SNA provides insights on the crucial players structuring the interplay of their team as well as the overall passing structure. Similarly, past performances of future opponents can be studied as part of the game preparation.

However, the information becomes more useful if put into context. Here, play-level analysis offers a more accurate breakdown of the emerging behavior in varying game situations facing different constraints. It implies an enhanced ability to control for situational variables and the connection to performance outcomes that describe interactions leading to success - or not. According to Praça et al. (2017), the resulting awareness of constraints and circumstances that lead to certain interactions can help in training. Here, certain constraints can be artificially reconstructed to produce desirable interaction patterns and, thus, use the knowledge for new training stimuli. As an example, imagine a setting in which the match analysis of an upcoming opponent reveals that certain playing formations or types of defense force the future opponent into specific passing patterns. Then, this knowledge can be used to reconstruct these setups in training and practice the anticipation and interception of passes by the defense. Here, MSTs could also assist in facilitating a comprehensible overview on the strongest passing connections within a team. After all, information presented to coaches and athletes should be intuitive and easy to comprehend (Glazier, 2010). Following the simplicity principle of centroids which represent the mean position of many players in a single variable, MSTs reduce the complexity of interactions by concentrating the necessary information on a graph with only a limited amount of edges (Passos et al., 2016).

The analysis can be even more enhanced by considering the sequential order of passing sequences at play-level as implemented in the third study of this dissertation. Among other advantages, it can produce insights on the role of players in passing sequences, e.g. the revelation of bridging players that connect team members during attacks. Thus far, metrics such as betweenness centrality were used to investigate the dependency on these types of players (Gonçalves et al., 2017). The insights are helpful for coaches in order to understand the vulnerability of the own or opponent team induced by the dependency on a focal player. Flow betweenness, as the novel and alternative measurement of betweenness, can more accurately highlight players that facilitate and control ball flow within their teams.

In general, the validity and, thus, usefulness of performance indicators generated from SNA heavily rely on their underlying normalization procedures. According to Sampaio and Leite (2013), this is crucial to conduct performance comparisons within and across matches and, thus, be able to objectively evaluate performances. For example, flow-based metrics are only useful in football if there is a relatively common basis of assessment in terms of the amount of plays. As the studies of this dissertation codified playing positions, substitutions lead to a simple reassignment of positions. This led to a common basis of plays to compare the involvement of playing positions in the interplay. However, tracking specific players in an actual match analysis might require further normalization or adjustments for playing time. Then, the proposed network metrics can also be useful in practice.

4.3 Limitations of Method

Despite the theoretical and practical impact, the application of SNA in performance analysis still faces some challenges and limitations that need to be addressed. Thus far, including this dissertation, there is a strict focus on the interactions of one team with limited consideration of the opponent. In team sports, the adversary team constraints the possible actions implying an impact on the opposing passing structure. Hence, disregarding the influence of the opponent team restricts a comprehensive analysis of the causes for certain passing behaviors. Apart from the consideration of attack formations in the second study on handball, this has not been implemented in this dissertation.

Moreover, this dissertation only focuses on successful passes between players. First, there is no further information provided on the quality of the pass. For example, McLean et al. (2017b) emphasize the degree of penetration as a quality feature of a pass by counting the number of overplayed opponents. Second, ties predominantly reflect passes between player to model the interactions within a team. This, of course, is an abstract and simplified view on the complex interactive behavior of sport teams. There might be other performance-relevant forms of communication between players, e.g. the actual verbal communication or joint defensive actions as proposed by Sasaki et al. (2017).

However, not only the modeling of ties faces limitations but also the specification of nodes. In this dissertation, nodes represented playing positions in order to model individual contribution while still being able to generalize and compare metrics across matches. Yet, playing positions are interpreted differently not only across but also within matches, which could impair the validity of the calculated metrics.

Lastly, while this dissertation focuses on the temporal aspects of interactions in team sports, the connection to spatial information in form of positional data is still limited. Incorporating information on the positioning of players of both teams at all times increases the understanding of the interaction context (Passos et al., 2016; Ramos et al., 2018). Thus, it can be a pivotal enhancement of the evaluation of interactions in team sports.

5 Outlook

Performance analysis in team sports remains a challenge due to the dynamic and complex interactive nature, especially in invasion games like football. SNA can contribute to a better understanding of the emerging interactive behavior under constraints. This dissertation underlines that this does not only apply to football but invasion games as such, provided an adequate adoption and modification of network procedures. Moreover, this dissertation breaks down interactions to a play-level analysis to account for variations in constraints and game situations as well as successful and unsuccessful performance outcomes. The consideration of the sequential order of passes contributes to the modeling of the dynamic nature of interactions in team sports. In summary, this dissertation contributes especially to the theoretical aspects of performance analysis which, in turn, can also have a substantial impact on practice enabled by a more precise modeling of team performance.

Future studies should focus on increasing the practical impact of SNA in team sports. Eventually, coaches and analysts should profit from the more adequate modeling of interactions and, thus, increased understanding of the performance of their team. This can be achieved by incorporating more situational variables at play-level and controlling for varying constraints such as relative positioning of the opposition. Moreover, the focus should be on the development of network metrics that actively model the sequential pattern of passes. While flow betweenness offers a first alternative to betweenness centrality, further traditional metrics need to be modified in order to consider the nature of ties in passing networks.

Besides, the modifications of the network approach in team sports might be transferable to other application areas beyond sports science. Indeed, there are cases where SNA researchers transferred knowledge from other areas to their field. Newman (2018) analyzed root networks of plants and found that the procedure could be transferred to networks in the field of geography, e.g. river networks. In team sports, passing networks are built from multiple separable sequences of interactions between players, following no particular route such as the shortest path. Traditional metrics facilitate a static perspective on the accumulated passing network across a certain time frame, while metrics such as flow betweenness consider the sequential order of passes during ball possessions. Hence, the application could be suitable for the analysis of networks that are equally built from many separable interaction or communication sequences.

6 Bibliography

Aquino, R., Carling, C., Palucci Vieira, L. H., Martins, G., Jabor, G., Machado, J., & Puggina, E. (2018). Influence of situational variables, team formation, and playing position on match running performance and social network analysis in Brazilian professional soccer players. J Strength Cond Res.

Araújo, D., & Davids, K. (2016). Team synergies in sport: theory and measures. Frontiers in psychology, 7, 1449.

Araujo, D., Davids, K., & Hristovski, R. (2006). The ecological dynamics of decision making in sport. Psychology of sport and exercise, 7(6), 653-676.

Araújo, D., Davids, K., & Passos, P. (2013). The intending–perceiving–acting cycle in sports performance. In Routledge Handbook of Sports Performance Analysis. Routledge, 50-59.

Arriaza-Ardiles, E., Martín-González, J. M., Zuniga, M. D., Sánchez-Flores, J., De Saa, Y., & García-Manso, J. M. (2018). Applying graphs and complex networks to football metric interpretation. Human movement science, 57, 236-243.

Atkin, R. (1974). Mathematical structure in human affairs. Heinemann Educational Publishers.

Balkundi, P., & Harrison, D. A. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. Academy of Management Journal, 49(1), 49-68.

Bastin, J., Craig, C., & Montagne, G. (2006). Prospective strategies underlie the control of interceptive actions. Human movement science, 25(6), 718-732.

Bavelas, A. (1950). Communication patterns in task-oriented groups. The Journal of the Acoustical Society of America, 22(6), 725-730.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. Journal of mathematical sociology, 2(1), 113-120.

Bonanno, G., Caldarelli, G., Lillo, F., & Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. Physical Review E, 68(4), 046130.

Borgatti, S. P. (2005). Centrality and network flow. Social networks, 27(1), 55-71.

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. Social networks, 28(4), 466-484.

Borgatti, S. P., & Halgin, D. S. (2011). On network theory. Organization science, 22(5), 1168-1181.

Borgatti, S. P., & Kidwell, V. L. (2011). Network Theory. The SAGE handbook of social network analysis. SAGE publications, 40-54.

Bourbousson, J., Poizat, G., Saury, J., & Seve, C. (2010). Team coordination in basketball: Description of the cognitive connections among teammates. Journal of Applied Sport Psychology, 22(2), 150-166.

Carling, C. W., Williams, A. M. & Reilly, T. (2005). The Handbook of Soccer Match Analysis. London: Routledge.

Castellano, J., & Echeazarra, I. (2019). Network-based centrality measures and physical demands in football regarding player position: Is there a connection? A preliminary study. Journal of sports sciences, 1-8.

Chaouachi, A., Brughelli, M., Levin, G., Boudhina, N. B. B., Cronin, J., & Chamari, K. (2009). Anthropometric, physiological and performance characteristics of elite team-handball players. Journal of sports sciences, 27(2), 151-157.

Clemente, F. M., Couceiro, M. S., Martins, F. M. L., & Mendes, R. S. (2014a). Using network metrics to investigate football team players' connections: A pilot study. Motriz: Revista de Educação Física, 20(3), 262-271.

Clemente, F. M., Martins, F. M. L., Couceiro, M. S., Mendes, R. S., & Figueiredo, A. J. (2014b). A network approach to characterize the teammates' interactions on football: A single match analysis. Cuadernos de Psicología del Deporte, 14(3), 141-148.

Clemente, F. M., Martins, F. M. L., Kalamaras, D., & Mendes, R. S. (2015a). Network analysis in basketball: inspecting the prominent players using centrality metrics. Journal of Physical Education and Sport, 15(2), 212.

Clemente, F. M., Martins, F. M. L., Wong, P. D., Kalamaras, D., Mendes, R. S. (2015b). Midfielder as the prominent participant in the building attack: A network analysis of national teams in FIFA World Cup 2014. International Journal of Performance Analysis in Sport, 15(2), 704-722.

Clemente, F. M., Martins, F. M. L., Kalamaras, D., Oliveira, J., Oliveira, P., & Mendes, R. S. (2015c). The social network analysis of Switzerland football team on FIFA World Cup 2014. Journal of Physical Education and Sport, 15(1), 136.

Clemente, F. M., Couceiro, M. S., Martins, F. M. L., & Mendes, R. S. (2015d). Using network metrics in soccer: a macro-analysis. Journal of human kinetics, 45(1), 123-134.

Clemente, F. M., Martins, F. M. L., & Mendes, R. S. (2016a). Analysis of scored and conceded goals by a football team throughout a season: a network analysis. Kinesiology: International journal of fundamental and applied kinesiology, 48(1), 103-114.

Clemente, F. M., José, F., Oliveira, N., Martins, F. M. L., Mendes, R. S., Figueiredo, A. J., Wong, D.P. & Kalamaras, D. (2016b). Network structure and centralization tendencies in professional football teams from Spanish La Liga and English Premier Leagues. Journal of Human Sport and Exercise, 11(3), 376-389.

Clemente, F. M., Silva, F., Martins, F. M. L., Kalamaras, D., & Mendes, R. S. (2016c). Performance Analysis Tool for network analysis on team sports: A case study of FIFA Soccer World Cup 2014. Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology, 230(3), 158-170.

Clemente, F. M., & Martins, F. M. L. (2017a). Who are the prominent players in the UEFA champions league? An approach based on network analysis. Walailak Journal of Science and Technology (WJST), 14(8), 627-636.

Clemente, F. M., & Martins, F. M. L. (2017b). Network structure of UEFA Champions League teams: association with classical notational variables and variance between different levels of success. International Journal of Computer Science in Sport, 16(1), 39-50.

Clemente, F. M. (2018). Performance outcomes and their associations with network measures during FIFA World Cup 2018. International Journal of Performance Analysis in Sport, 18(6), 1010-1023.

Cotta, C., Mora, A. M., Merelo, J. J., & Merelo-Molina, C. (2013). A network analysis of the 2010 FIFA world cup champion team play. Journal of Systems Science and Complexity, 26(1), 21-42.

Cousens, L., & Slack, T. (1996). Emerging patterns of inter-organizational relations: A network perspective of North American professional sport leagues. European Journal for Sport Management, 3, 48-69.

Davids, K., Araújo, D., & Shuttleworth, R. (2005). Applications of dynamical systems theory to football. Science and football V, 537-550.

Davids, K., Hristovski, R., Araújo, D., Serre, N. B., Button, C., & Passos, P. (2013). Complex systems in sport. Routledge.

Dey, P., Ganguly, M., & Roy, S. (2017). Network centrality based team formation: A case study on T-20 cricket. Applied computing and informatics, 13(2), 161-168.

Dong, J., & Horvath, S. (2007). Understanding network concepts in modules. BMC systems biology, 1(1), 24.

Duarte, R., Araújo, D., Correia, V., & Davids, K. (2012). Sports teams as superorganisms. Sports medicine, 42(8), 633-642.

Duarte, R., Araújo, D., Correia, V., Davids, K., Marques, P., & Richardson, M. J. (2013). Competing together: Assessing the dynamics of team–team and player–team synchrony in professional association football. Human Movement Science, 32(4), 555-566.

Duch, J., Waitzman, J. S., & Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. PloS one, 5(6), e10937.

Fagiolo, G. (2007). Clustering in complex directed networks. Physical Review E, 76(2), 026107.

Fewell, J. H., Armbruster, D., Ingraham, J., Petersen, A., & Waters, J. S. (2012). Basketball teams as strategic networks. PloS one, 7(11), e47445.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. Sociometry, 35-41.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social networks, 1(3), 215-239.

Frencken, W., Lemmink, K., Delleman, N., & Visscher, C. (2011). Oscillations of centroid position and surface area of soccer teams in small-sided games. European Journal of Sport Science, 11(4), 215-223.

Gama, J., Couceiro, M., Dias, G., & Vaz, V. (2015). Small-world networks in professional football: conceptual model and data. European Journal of Human Movement, 35, 85-113.

Gama, J., Passos, P., Davids, K., Relvas, H., Ribeiro, J., Vaz, V., & Dias, G. (2014). Network analysis and intra-team activity in attacking phases of professional football. International Journal of Performance Analysis in Sport, 14(3), 692-708.

Glazier, P. S. (2010). Game, set and match? Substantive issues and future directions in performance analysis. Sports medicine, 40(8), 625-634.

Gloor, P., Paasivaara, M., Schoder, D., & Willems, P. (2006). Correlating performance with social network structure through teaching social network analysis. In Working Conference on Virtual Enterprises.Springer, 265-272.

Gómez, M. Á., Lago-Peñas, C., & Pollard, R. (2013). Situational variables. In Routledge handbook of sports performance analysis. Routledge, 277-287.

Gonçalves, B., Coutinho, D., Santos, S., Lago-Penas, C., Jiménez, S., & Sampaio, J. (2017). Exploring team passing networks and player movement dynamics in youth association football. PloS one, 12(1), e0171156.

Granovetter, M. S. (1977). The strength of weak ties. In Social networks (pp. 347-367). Academic Press.

Gréhaigne, J. F., & Godbout, P. (1995). Tactical knowledge in team sports from a constructivist and cognitivist perspective. Quest, 47(4), 490-505.

Grund, T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. Social Networks, 34(4), 682-690.

Hohmann, A., & Brack, R. (1983). Theoretische Aspekte der Leistungsdiagnostik im Sportspiel. Leistungssport, 13(2), 5-10.

Hohmann, A., Lames, M., & Letzelter, M. (2002). Einführung in die Trainingswissenschaft. Wiebelsheim.

Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. Journal of sports sciences, 20(10), 739-754.

Hughes, M., & Franks, I. (2007). The essentials of performance analysis: an introduction. Routledge.

Jackson, M. O. (2010). Social and economic networks. Princeton university press.

Korte, F., & Lames, M. (in press). Passing network analysis of positional attack formations in handball. Journal of Human Kinetics.

Korte, F., & Lames, M. (2018). Characterizing different team sports using network analysis. Current Issues in Sport Science, 3:005.

Korte F., Link D., Groll J., & Lames M. (2019). Play-by-Play Network Analysis in Football. Front. Psychol. 10:1738.

Lai, M., Meo, R., Schifanella, R., & Sulis, E. (2018). The role of the network of matches on predicting success in table tennis. Journal of sports sciences, 36(23), 2691-2698.

Lames, M., & Hansen, G. (2001). Designing observational systems to support top-level teams in game sports. International journal of performance analysis in sport, 1(1), 83-90.

Lames, M., & McGarry, T. (2007). On the search for reliable performance indicators in game sports. International Journal of Performance Analysis in Sport, 7(1), 62-79.

Lames, M. (2018). Chance involvement in goal scoring in football–an empirical approach. German Journal of Exercise and Sport Research, 48(2), 278-286.

Lemmink, K. A., & Frencken, W. (2013). Tactical performance analysis in invasion games. In Routledge handbook of sports performance analysis. Routledge, 89-100.

Letzelter, H., & Letzelter, M. (1982). Die Struktur sportlicher Leistungen als Gegenstand der Leistungsdiagnostik in der Trainingswissenschaft. Leistungssport, 12(5), 351-361.

Maslovat, D., & Franks, I. M. (2008). The need for feedback. The essentials of performance analysis: An introduction, 1-7.

Mateus, J. (2005). In pursuit of an ecological and fractal approach to football coaching. Science and Football V, 571-583.

McGarry, T. (2009). Applied and theoretical perspectives of performance analysis in sport: Scientific issues and challenges. International Journal of Performance Analysis in Sport, 9(1), 128-140.

McGarry, T., Anderson, D. I., Wallace, S. A., Hughes, M. D., & Franks, I. M. (2002). Sport competition as a dynamical self-organizing system. Journal of sports sciences, 20(10), 771-781.

McLean, S., Salmon, P. M., Gorman, A. D., Naughton, M., & Solomon, C. (2017a). Do inter-continental playing styles exist? Using social network analysis to compare goals from the 2016 EURO and COPA football tournaments knock-out stages. Theoretical Issues in Ergonomics Science, 18(4), 370-383.

McLean, S., Salmon, P. M., Gorman, A. D., Read, G. J., & Solomon, C. (2017b). What's in a game? A systems approach to enhancing performance analysis in football. PloS one, 12(2), e0172565.

McLean, S., Salmon, P. M., Gorman, A. D., Stevens, N. J., & Solomon, C. (2018). A social network analysis of the goal scoring passing networks of the 2016 European Football Championships. Human movement science, 57, 400-408.

McLean, S., Salmon, P. M., Gorman, A. D., Dodd, K., & Solomon, C. (2019). Integrating communication and passing networks in football using social network analysis. Science and Medicine in Football, 3(1), 29-35.

Moreno, J. L. (1934). Nervous and mental disease monograph series, no 58. Who shall survive?: A new approach to the problem of human interrelations. Washington, DC, US: Nervous and Mental Disease Publishing Co.

Nadel, S. F. (1957). The theory of social structure. Glencoe, IL. Free Press.

Newell, K. M. (1986). Constraints on the development of coordination. In Wade, M. G., Whiting, H. T. A., Motor Development in Children: Aspects of Coordination and Control, 341-360.

Newman, M. E. (2001). The structure of scientific collaboration networks. Proceedings of the national academy of sciences, 98(2), 404-409.

Newman, M. E. (2016). Mathematics of networks. The new Palgrave dictionary of economics, 1-8.

Newman, M. E. (2018). Networks. Oxford university press.

Oliveira, P., & Clemente, F. M. (2018). Network properties and performance variables and their relationships with distance covered during elite soccer games. Journal of Physical Education and Sport, 18, 1045-1049.

Padgett, J. F., & Ansell, C. K. (1993). Robust Action and the Rise of the Medici, 1400-1434. American journal of sociology, 98(6), 1259-1319.

Passos, P., Araújo, D., & Volossovitch, A. (2016). Performance analysis in team sports. Taylor Francis.

Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., & Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. Journal of Science and Medicine in Sport, 14(2), 170-176.

Peixoto, D., Praça, G. M., Bredt, S., & Clemente, F. M. (2017). Comparison of network processes between successful and unsuccessful offensive sequences in elite soccer. Human Movement, 18(5), 48-54.

Pena, J. L., & Touchette, H. (2012). A network theory analysis of football strategies. arXiv preprint arXiv:1206.6904.

Pina, T. J., Paulo, A., & Araújo, D. (2017). Network characteristics of successful performance in association football. A study on the UEFA Champions League. Frontiers in Psychology, 8, 1173.

Pollard, R., & Reep, C. (1997). Measuring the effectiveness of playing strategies at soccer. Journal of the Royal Statistical Society: Series D (The Statistician), 46(4), 541-550.

Praça G. M., Clemente, F. M., Bredt, S. D. G. T., Chagas, M. H., Morales, J. C. P., Peixoto, G. H. D. C., & Greco, P. J. (2018). Analysis of network properties and tactical behavior of u-17 soccer athletes with different tactical skills. Journal of Physical Education, 29.

Praça, G. M., Clemente, F. M., Pereira de Andrade, A. G., Perez Morales, J. C., & Greco, P. J. (2017). Network analysis in small-sided and conditioned soccer games: The influence of additional players and playing position. Kinesiology: International journal of fundamental and applied kinesiology, 49(2), 185-193.

Praça, G. M., Lima, B. B., Bredt, S. D. G. T., Clemente, F. M., & Andrade, A. G. (2019). Influence of match status on players' prominence and teams' network properties during 2018 FIFA World Cup knockout phase. Frontiers in psychology, 10, 695.

Proctor, C. H., & Loomis, C. P. (1951). Analysis of sociometric data. Research methods in social relations, 2, 561-585.

Radicchi, F. (2011). Who is the best player ever? A complex network analysis of the history of professional tennis. PloS one, 6(2), e17249.

Ramos, J., Lopes, R. J., & Araújo, D. (2018). What's next in complex networks? Capturing the concept of attacking play in invasive team sports. Sports medicine, 48(1), 17-28.

Ratten, V., Ratten, H., & Cobbs, J. B. (2011). The dynamics of relationship marketing in international sponsorship networks. Journal of Business & Industrial Marketing.

Ribeiro, J., Silva, P., Duarte, R., Davids, K., & Garganta, J. (2017). Team sports performance analysed through the lens of social network theory: implications for research and practice. Sports medicine, 47(9), 1689-1696.

Roethlisberger, F. J., & Dickson, W. J. (1939). Management and the Worker Harvard University Press. Cambridge, Mass.

Sampaio, J., & Leite, N. (2013). Performance indicators in game sports. In Routledge handbook of sports performance analysis. Routledge, 133-144.

Sampaio, J., & Maçãs, V. (2012). Measuring tactical behaviour in football. International journal of sports medicine, 33(05), 395-401.

Sarmento, H., Clemente, F. M., Araújo, D., Davids, K., McRobert, A., & Figueiredo, A. (2018). What performance analysts need to know about research trends in association football (2012–2016): A systematic review. Sports medicine, 48(4), 799-836.

Sasaki, K., Yamamoto, T., Miyao, M., Katsuta, T., & Kono, I. (2017). Network centrality analysis to determine the tactical leader of a sports team. International Journal of Performance Analysis in Sport, 17(6), 822-831.

Scott, J. (1988). Social network analysis. Sociology, 22(1), 109-127.

Seevers, M. T., Skinner, S. J., & Dahlstrom, R. (2010). Performance implications of a retail purchasing network: The role of social capital. Journal of Retailing, 86(4), 310-321.

Shaw, M. E. (1954). Group structure and the behavior of individuals in small groups. The Journal of psychology, 38(1), 139-149.

Snijders, T. A. (1981). The degree variance: an index of graph heterogeneity. Social networks, 3(3), 163-174.

Tenga, A., Ronglan, L. T., & Bahr, R. (2010). Measuring the effectiveness of offensive match-play in professional soccer. European Journal of Sport Science, 10(4), 269-277.

Travassos, B., Araújo, D., Vilar, L., & McGarry, T. (2011). Interpersonal coordination and ball dynamics in futsal (indoor football). Human Movement Science, 30(6), 1245-1259.

Trequattrini, R., Lombardi, R., & Battista, M. (2015). Network analysis and football team performance: a first application. Team Performance Management, 21(1/2), 85-110.

Volossovitch, A. (2013). Handball. In Routledge handbook of sports performance analysis. Routledge, 398-410.

Vuleta, D., Milanović, D., & Sertić, H. (2003). Relations among variables of shooting for a goal and outcomes of the 2000 Men's Handball Championship matches. Kinesiology: international journal of fundamental and applied kinesiology, 35(2), 168-183.

Wagner, H., Finkenzeller, T., Würth, S., & Von Duvillard, S. P. (2014). Individual and team performance in team-handball: A review. Journal of sports science & medicine, 13(4), 808.

Wäsche, H., Dickson, G., Woll, A., & Brandes, U. (2017). Social network analysis in sport research: an emerging paradigm. European Journal for Sport and Society, 14(2), 138-165.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications (Vol. 8). Cambridge university press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. nature, 393(6684), 440.

Yamamoto, Y. (2010). Scale-free property of the passing behaviour in a team sport. International Journal of Sport and Health Science, 1002250049-1002250049.

A Appendix

A.1 Study 1 - Permission for the inclusion in dissertation

Wednesday, October 3, 2018 at 11:04:32 AM Pacific Daylight Time

Betreff: WG: CISS - Artikel online
Datum: Mittwoch, 3. Oktober 2018 um 1:35:49 AM Nordamerikanische Westküsten-Sommerzeit
Von: Kopp, Martin
An: Korte, Florian
CC: Holzer, Daniela
Anlagen: image001.jpg

Lieber Herr Korte, nach den Regularien der Open-Access Policy, können Sie Ihren Artikel sehr gerne in Ihre Dissertation einbinden.

Mit freundlichen Grüßen, Martin Kopp

universität innsbruck
Institut für Sportwissenschaft

Univ.-Prof. Dr. Martin Kopp
Professur für Sportpsychologie und Institutsleitung
Fürstenweg 185
A-6020 Innsbruck, Österreich
Tel: (43)- 512-507-45851
FAX: (43)-512-507-45999
E-mail: Martin.Kopp@uibk.ac.at
http://www.uibk.ac.at/isw/
https://tinyurl.com/y995m9mb

Diese E-Mail wurde mit einem Spracherkennungsprogramm erstellt. Geringfügige Fehler können trotz Kontrolle auftreten. Danke für Ihr Verständnis.

Von: Korte, Florian <korte@cdtm.de>
Gesendet: Dienstag, 2. Oktober 2018 22:02
An: Holzer, Daniela <D.Holzer@uibk.ac.at>
Betreff: Re: CISS - Artikel online

Liebe Frau Holzer,

ich wollte mich kurz erkundigen, ob der Prozess der Zahlung der Rechnung über die TU München erfolgreich abgewickelt wurde?

Darüber hinaus möchte ich kurz final sichergehen, dass ich den Artikel, nach den Regularien der Open-Access Policy, in meine Dissertation einbinden darf?

Besten Dank vorab für Ihre Rückmeldung.

Gruß,
Florian Korte

A.2 Study 1 - Original article

Current Issues in Sport Science 3 (2018)

Characterizing different team sports using network analysis

Florian Korte[1, 2, *] & Martin Lames[2]

1 Center for Digital Technology and Management, Munich, Germany
2 Department of Sport and Health Sciences, Technical University of Munich, Germany
* Corresponding author: Center for Digital Technology and Management, Marsstraße 20-22, 80335 Munich, Germany,
Tel: +49 (0) 89-28928163, Fax: +49 (0) 89-28928459
Email: korte@cdtm.de

COMMENTARY

Article History:
Submitted 12th October 2017
Accepted 14th March 2018
Published 25th April 2018

Handling Editor:
Ernst-Joachim Hossner
University of Bern, Switzerland

Editor-in-Chief:
Martin Kopp
University of Innsbruck, Austria

Reviewers:
Reviewer 1: Filipe Manuel Clement
Instituto Politecnico de Viana do Castelo, Portugal
Reviewer 2: Anonymous

ABSTRACT

Team sports are complex dynamic systems based on the frequent interaction of various players. Recently, social network analysis has been introduced to the study of sports dynamics in order to quantify the involvement of individual players in the interplay and to characterize the organizational processes used by teams. Nonetheless, only a limited set of team sports has been assessed to date, and the focus of most studies has been on the application of small sets of network metrics to a single sport. Our study aims at comparing the network patterns of different team sports in order to contribute to the understanding of their underlying nature. It considers three invasion games, namely professional matches from basketball, football and handball. By applying relevant centrality measures and minimum spanning trees a first comparison between the nature of interplay in various team sports is offered as well as a deeper understanding of the role of different tactical positions in each sport. The point guard in basketball, defensive midfielder in football and center in handball are identified as the most central tactical positions. Direct interplay is most balanced in football followed by basketball and handball. A visualization of the basic structure of interplay for each sport is achieved through minimum spanning trees.

Keywords:
social network analysis – team sports – interaction matrices – minimum spanning trees

Citation:
Korte, F., & Lames, M. (2018). Characterizing different team sports using network analysis. *Current Issues in Sport Science, 3:005*. doi: 10.15203/CISS_2018.005

Introduction

Matches, or games, in team sports can be seen as complex dynamic systems (Glazier & Davids, 2009). The frequent interaction of various players is an integral part of any team sports match (Passos, Araújo & Volossovitch, 2016). Hence, a team must be regarded as more than the sum of its parts, and the secret to successful performance is believed to lie in the collective action of team members (Grund, 2012). Understanding the patterns of play is important to deduce the nature of the sport. Moreover, the individual contribution of each player to the organizational process is highly relevant to revealing how a team functions (Vilar, Araújo, Davids, & Bar-Yam, 2013). The complexity of matches and team dynamics makes breaking down such patterns difficult, creating an ongoing challenge for performance analysis in team sports.
There is an increasing interest in applying Social Network Analysis (SNA), a method that exploits familiar performance variables such as passes, in order to detect patterns in the interplay of teams (Clemente & Martins, 2017). Network approaches focus on breaking down the web of interactions in systems of multiple agents also referred to as nodes (Passos et al., 2011). Traditional application areas of this method can be found in biological (e.g. spread of diseases) and sociological (e.g. acquaintance net-

2018 | *innsbruck* university press, Innsbruck
Current Issues in Sport Science | ISSN 2414-6641 | http://www.ciss-journal.org/
Vol. 3 | DOI 10.15203/CISS_2018.005

works) contexts. In sports, the frequent interaction between a limited set of players, e.g. through the passing of a ball qualifies network theory as a powerful performance analysis method.
Clemente, Martins, Wong, Kalamaras and Mendes (2015b) analyze professional football matches by applying SNA. On a micro level, i.e. focusing on the prominence of individual players in a team, the authors identify the position of the central midfielder as the most prominent player in their study, as midfielders are responsible for building offensive lines of attack. Pena and Touchette (2012) detect certain cliques within football teams that interact more frequently than others. This is in line with another micro-level study by Gama et al. (2014), who find that only a subset of players in football teams is responsible for the majority of interaction and thus shaping the pattern of play. On both, micro and macro level, i.e. focusing on the collective organization of a team, Duch, Waitzman and Amaral (2010) identify a strong connection between several network measures and traditional performance indicators whereas Grund (2012) connects the distribution of individual networks measures to performance outcomes. In his macro-level analysis, the author finds that successful teams in football demonstrate a more balanced interplay.
In basketball, SNA has been applied in professional and amateur settings. Fewell, Armbruster, Ingraham, Petersen and Waters (2012) and Clemente, Martins, Kalamaras and Mendes (2015a) identify the Point Guard as the dominant player structuring plays for the team.
However, the set of sports that SNA has been applied to has been limited so far. Moreover, the focus of most studies has been on the application of small sets of network metrics to a single sport. Our study aims at comparing the network patterns of different team sports in order to contribute to the understanding of their underlying nature. It considers three invasion games, namely professional matches from basketball, football and handball. The overarching task of each team trying to collectively outperform or -score its opponent unites these popular team sports. However, as they differ in their environmental constraints (e.g. areas, rules), different interaction patterns are needed in order to succeed (Araújo & Davids, 2016).
SNA enables us to investigate the resulting complex webs of interaction between the players in the different sports. To ensure a thorough analysis, individual and team metrics are applied alongside the computation of minimum spanning trees, a network technique that facilitates an intuitive visualization of the strongest relationships in complex networks revealing the basic structure of the sports.
In combination with the macro-level analysis, i.e. applying team metrics, this assesses the overall interaction patterns. The micro-level analysis, i.e. applying individual metrics, is specifically targeted at revealing the dominant tactical positions in terms of their involvement in the interplay for each sport and who are responsible for structuring these patterns. The combined analysis enables us to break down the complex organizational processes within teams and thus contributing to the understanding of the underlying nature of basketball, football and handball.
To our knowledge, this is the first study that attempts a comparison of different team sports applying SNA. Furthermore, it is the first analysis that takes handball into consideration along with football and basketball and applies minimum spanning trees in the context of team sports.
Hence, this study breaks down the underlying complexity of team sports by characterizing and quantifying individual and team performance through SNA.

Methods

Samples

For each sport, eight knockout round matches in the men´s competition at major professional tournaments are considered for analysis, minimizing the home/away bias (Courneya & Carron, 1992). For basketball and handball the knockout stages at the Rio 2016 Summer Games Olympics tournaments are recorded and analyzed. For football, the authors consider the last eight matches from the knockout stage of the FIFA World Cup 2014 tournament. A total of 16 adjacency matrices for each sport are generated, capturing the interaction between players of each team. A total of 4059 passes are analyzed in basketball, 6934 in football and 8054 in handball.

Procedure

In order to apply SNA, adjacency matrices capture the passing distribution seen in every analyzed match. The matrices are constructed from a set of nodes and edges for every team respectively. Players represent nodes such that the number of passes between them defines the edge weight. The overall match-based interaction matrix per team is a result of an aggregation of the units of attack defined as the moment from ball recovery until possession is lost (Passos et al., 2011).
The tracking process for basketball and handball games was executed through video analysis applying the software Dartfish®. The passing distribution at the FIFA World Cup 2014 tournament was provided in the official FIFA match reports on their website (www.fifa.com/worldcup/archive/brazil2014). In a thorough post-match analysis players were assigned to their respective tactical position to ensure the comparability between teams and focus on the tactical aspects of each sport. In line with O`Donoghue (2009), we acknowledge the increasing complexity of tactical roles in team sports, i.e. forwards taking on defending tasks in football. Players might temporarily occupy different areas on the pitch and fulfill different tasks which can be acknowledged as part of the role repertoire of the different tactical positions, especially in football. Eventually, this is part of why we see complex webs of interaction in team sports and why we expect that this finds its expression in the results of our analysis. The definition of tactical roles for the three sports is displayed in Table 1.

Table 1: List of tactical roles in basketball, football and handball

Basketball	Football		Handball
Center (C)	Defensive Midfielder (DM)	Offensive Midfielder (OM)	Center (C)
Point Guard (PG)	Goalkeeper (GK)	Right Central Defender (RCD)	Left Back (LB)
Power Forward (PF)	Left Central Defender (LCD)	Right Defender (RD)	Left Wing (LW)
Shooting Guard (SG)	Left Defender (LD)	Right Forward (RF)	Pivot (P)
Small Forward (SF)	Left Forward (LF)	Right Midfielder (RM)	Right Back (RB)
	Left Midfielder (LM)		Right Wing (RW)

Following the codification for each tactical position ensured that frequent substitutions of players lead to a reassignment of the given tactical positions. Predominantly, substitutions lead to a direct replacement for the corresponding tactical position, meaning the player who was codified to a specific position was replaced by his substitute. However, substitutions occasionally implied the reassignment on multiple positions, mostly in basketball and handball. To detect these changes, each unit of attack was considered separately. Tracking and codification processes were executed by researchers with more than ten years of experience in the sports described. In order to ensure the reliability of the study, Cohen´s kappa and Gwet's AC1 inter-rater statistic were computed in a two-stage process (Gwet, 2001). In a first step, the agreement on the occurrence of passes was analyzed using Gwet`s statistics. In a second step, the agreement on passer and pass receiver was tested applying Cohen's Kappa. 12.5% of the overall data were tested for reliability purposes. The Kappa (Gwet, 2001) values were above 0.94 (0.85) respectively for each sport, ensuring the reliability of the data.

Network Metrics

For the 16 adjacency matrices in each sport a set of individual- and team-related centrality network metrics are computed. The analysis was carried out using the software Matlab® and the visualization of networks was generated by applying Cytoscape®.

Centrality calculations allow a quantification of the influence of tactical positions on their team´s interplay as well as the balance of influence between players overall. To account for the nature of the sports, metrics that consider weighted directed graphs were applied. This allows for a breakdown of the connection between any two players in both passing directions. For individual (or micro-level) analysis weighted in-/out-degree, weighted betweenness and weighted closeness were computed. For team (or macro-level) analysis, the corresponding centralization values were calculated. These metrics are explained in detail in the following.

Individual Metrics Weighted in-degree (C_{WID}), also referred to as Prestige, is the sum of the incoming weighted edge values of a node. Hence, these metrics capture the number of successfully received passes of a player and a high value is often taken as a first indicator for the prominence of a particular player (Clemente et al., 2015b). Team members appear to trust this player, when in possession, to positively contribute to the team´s performance and therefore target him more frequently than others. Weighted out-degree (C_{WOD}), also referred to as Centrality, is the sum of outgoing weighted edges of a node. In the context of sports, (C_{WOD}) is the number of completed passes of a player and a high value is often associated with a high contribution to ball circulation (Clemente et al., 2015b).

We also calculate the ratio C_{WID} / C_{WOD} to assess a potential deviation between the share in pass reception and execution. A player with a higher reception than execution share, i.e. a value above 1, could indicate a player who rather finishes attacks. He frequently receives the ball from team members to execute shots on target rather than passing on. The opposite, i.e. a value below 1, might be a player who initiates attacks.

Weighted betweenness (C_{WB}) assesses how often a node is on the shortest path between two other nodes (Wassermann & Faust, 1994). A modified version of the standard computation of C_{WB} according to Newman (2001) is applied, which is more suitable for team sports since it favors strong connections rather than penalizing them. It measures how often a player is in between the most frequent passing connections of any other two players, thus functioning as a bridging unit (Pena & Touchette, 2012). As this implies a certain level of dependency on that particular player to ensure ball circulation it can be considered as a playmaker indicator.

Weighted closeness (C_{WC}) addresses how well connected a node is to all other nodes, directly or indirectly, within a network following Freeman (1978) and Opsahl, Agneessens and Skvoretz (2010). In a nearly complete network, i.e. in which almost every node is connected to each other, the metric can be seen as a more sophisticated approach to the weighted degree computations as the distribution of weights between other nodes is taken into account. In team sports, C_{WC} describes the how well a player directly or indirectly interacts with all other team members on the field. Hence, a player with high weighted degree values but comparatively low weighted closeness value might only interact strongly with a subset of his team members.

Team Metrics Centralization measures are concerned with the distribution of the individual metrics in a network. Following Freeman (1978) and Wasserman and Faust (1994), weighted indegree centralization (C_{WIDC}) captures the deviations from all indegree values to the highest value in the network adjusted by the number of passes and the number of players. This adjustment in the computation allows a comparison between different sports. Weighted out-degree centralization (C_{WODC}), weighted betweenness centralization (C_{WBC}) and weighted closeness centralization (C_{WCC}) is calculated accordingly.

By construction, all centralization values are bounded between 0 and 1. A network is regarded as highly centralized, i.e. a value close to 1, when the score of a particular node clearly outweighs the scores of all others and rather decentralized, i.e. a value close to 0, when the scores are similar among all nodes (Grund, 2012). In a sports context, C_{WIDC} and C_{WODC} scores can be seen as indicators for the balance of direct interplay in a team. C_{WBC} and C_{WCC} scores signal how balanced the influence on the overall interplay is within the team, considering direct and indirect connections. In general, high values could imply that interplay depends on only a subset of players.

For reasons of comparability between different matches, we normalized all centrality values by the total scores of the respective metrics following Leydesdorff (2007). The values themselves have no direct relevance. Relative comparisons between the different values of a respective metric for the tactical positions were highly crucial.

Visualization

A more intuitive visualization of the underlying structure of the networks was allowed for by computing minimum spanning trees (MSTs) for each sport. MSTs are meant to provide a revelation of the strongest relationships in complex networks (Mantegna, 1999). As a visualization method, they reduce the complexity of connected graphs of n nodes with up to n(n-1) connections to the strongest n-1 edges under the side condition that each node is still contained. According to Araújo and Davids (2016), sport teams demonstrate a task-specific organization to reach a common goal under certain constraints. In past studies, MSTs have been applied to visualize how sets of team members organize themselves to form an effective collective organization for a specific task (Lappas, Liu, & Terzi, 2009; Li & Shan, 2010). Hence, we apply MSTs to trace how teams consisting of a limited set of players organize their interplay in order to achieve group success. The method reduces the complex network of passes to the most basic structure presenting the most intensive connections under the consideration of all players.

As MSTs are only applicable to undirected graphs, the total passing intensity between pairs of players is considered in their construction. Reducing the amount of edges and thus complexity of the otherwise nearly complete networks, offers an alternative perspective on the pattern of interplay of the different team sports and hierarchical structure of weighted graphs (Gower & Ross, 1969).

Statistical Procedures

The authors of this paper utilized multiple one-way ANOVA to test for statistical differences between the centrality levels of the tactical positions within each sport, and between the analyzed sports. The assumption of normality for dependent variables was tested using Kolmogorov-Smirnov tests (p-value < .05). The assumption of homogeneity for groups' variances was examined by using Levene's test. There were no violations of either normality or homogeneity. Pairwise comparisons were established by running Bonferroni post-hoc tests. The statistical analyses were all conducted at a significance level of $p < .05$ using Matlab®. Following Ferguson (2009) and Clemente and Martins (2017), η^2 is reported to interpret the effect size according to the following criteria: no effect ($\eta^2 < .04$); small effect ($.04 \leq \eta^2 < .25$); moderate effect ($.25 \leq \eta^2 < .64$); strong effect ($\eta^2 \geq .64$).

Results

The tests found statistical differences in the dependent variables for all centrality measures applied for the three team sports considered in this study. The η^2 values reported in Table 2 almost all demonstrate moderate to strong effects sizes for the multiple one-way ANOVA in this study.

Individual Parameters

Table 3 shows the descriptive statistics and post-hoc results for tactical positions in basketball. The PG position is assigned the highest values for all centrality metrics and is significantly more central than every other tactical position. For weighted betweenness, the normalized value of the PG is 0.87 and thus more than ten times higher than the next ranked tactical position. There is no value assigned here for the forward positions implying that no strongest connection between any two players on the team runs via those tactical positions. In general, the other four tactical positions demonstrate similar values and no statistical differences are found between them for the other metrics applied in this study.

The C_{WID} / C_{WOD} ratios are shown in Figure 1. Notable in the ratio revealed is the relatively low value for the center position. Here, the share in pass completion rate outweighs the share in pass reception.

Table 2: Effect size values η^2 for multiple one-way ANOVA

	Basketball	Football	Handball		All
C_{WID}	.59 (moderate)	.23 (small)	.92 (strong)	C_{WIDC}	.89 (strong)
C_{WOD}	.46 (moderate)	.27 (moderate)	.92 (strong)	C_{WODC}	.81 (strong)
C_{WB}	.87 (strong)	.32 (moderate)	.91 (strong)	C_{WBC}	.89 (strong)
C_{WC}	.72 (strong)	.44 (moderate)	.93 (strong)	C_{WCC}	.83 (strong)

No effect ($\eta^2 < .04$); small effect ($0.04 \le \eta^2 < .25$); moderate effect ($.25 \le \eta^2 < .64$); strong effect ($\eta^2 \ge .64$)

Table 3: Descriptive statistics and post-hoc results for basketball

	PG	SG	SF	PF	C
C_{WID}	$0.30\ (0.04)_{all}$	$0.20\ (0.03)_{PG}$	$0.16\ (0.03)_{PG}$	$0.16\ (0.02)_{PG}$	$0.17\ (0.02)_{PG}$
C_{WOD}	$0.28\ (0.04)_{all}$	$0.18\ (0.04)_{PG}$	$0.17\ (0.03)_{PG}$	$0.16\ (0.03)_{PG}$	$0.21\ (0.03)_{PG}$
C_{WB}	$0.87\ (0.20)_{all}$	$0.07\ (0.19)_{PG}$	-	-	$0.05\ (0.10)_{PG}$
C_{WC}	$0.27\ (0.02)_{all}$	$0.19\ (0.03)_{PG}$	$0.17\ (0.02)_{PG}$	$0.17\ (0.02)_{PG}$	$0.19\ (0.02)_{PG}$

Subscripts indicate to which tactical positions given value is statistically different for $p < .05$, e.g. *PG*: given value is statistically different to the value of the point guard; *All:* value is statistically different to all other tactical positions.

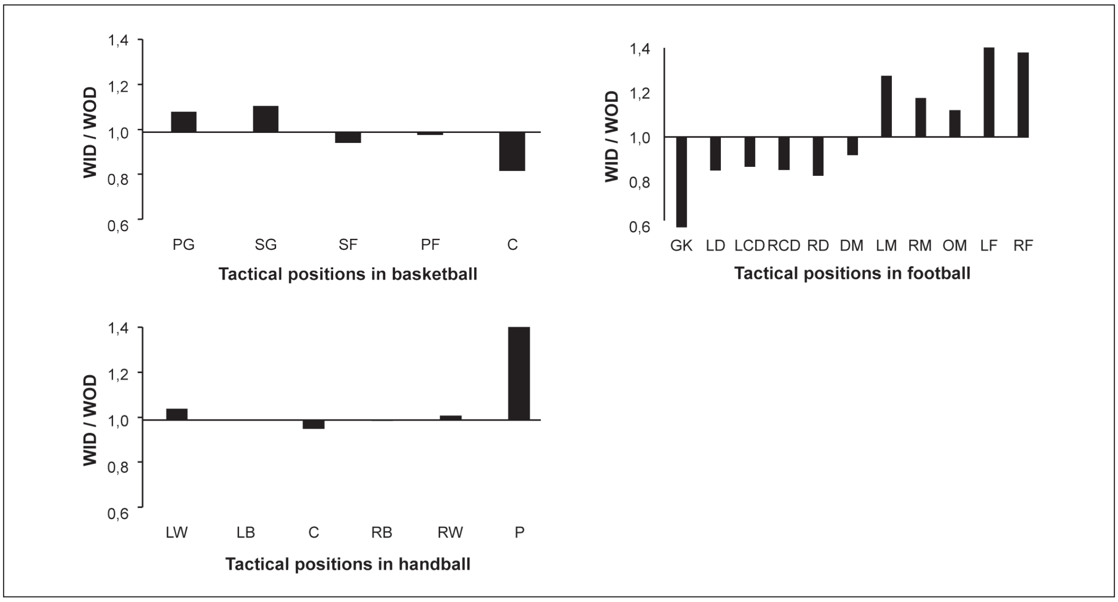

Figure 1: WID/WOD ratios for basketball, football and handball

Table 4: Descriptive statistics and post-hoc results for football

	GK	LD	LCD	RCD	RD	DM	LM	RM	OM	LF	RF
C_{WID}	0.03 $(0.01)_{all}$	0.08 $(0.02)_{GK}$	0.09 $(0.02)_{GK}$	0.09 $(0.02)_{GK}$	0.10 $(0.02)_{GK}$	0.12 $(0.03)_{GK}$	0.10 $(0.02)_{GK}$	0.11 $(0.02)_{GK}$	0.11 $(0.02)_{GK}$	0.09 $(0.04)_{GK}$	0.09 $(0.02)_{GK}$
C_{WOD}	0.06 $(0.02)_{mult}$	0.10 $(0.02)_{mult}$	0.10 $(0.02)_{mult}$	0.11 $(0.02)_{mult.}$	0.12 $(0.02)_{mult}$	0.13 $(0.02)_{GK,LM,Fs}$	0.08 $(0.02)_{mult}$	0.09 $(0.02)_{mult}$	0.10 $(0.02)_{mult}$	0.06 $(0.03)_{mult}$	0.06 $(0.02)_{mult}$
C_{WB}	0.00 $(0.01)_{mult}$	0.08 $(0.09)_{mult}$	0.12 $(0.06)_{mult}$	0.13 $(0.07)_{mult}$	0.18 $(0.14)_{mult}$	0.18 $(0.10)_{all-CDs,OM,RD}$	0.05 $(0.05)_{mult}$	0.07 $(0.07)_{mult}$	0.11 $(0.09)_{mult}$	0.05 $(0.08)_{mult}$	0.03 $(0.06)_{mult}$
C_{WC}	0.06 $(0.01)_{all}$	0.09 $(0.01)_{mult}$	0.09 $(0.01)_{mult}$	0.10 $(0.01)_{mult}$	0.10 $(0.01)_{mult}$	0.11 $(0.01)_{GK,Fs,LD,LF,LM}$	0.09 $(0.01)_{mult}$	0.10 $(0.01)_{mult}$	0.10 $(0.01)_{mult}$	0.08 $(0.02)_{mult}$	0.08 $(0.01)_{mult}$

Subscripts indicate to which tactical positions given value is statistically different for $p < .05$, e.g. *GK*: given value is statistically different to the value of the goalkeeper; *All:* value is statistically different to all other tactical positions; *All-"tactical position(s)"*: value is statistically different to all other tactical positions *except* the listed ones; *Mult*: value is statistically different to various tactical positions that are not part of further analysis in this study; *Fs* includes *LF* and *RF*; CDs includes *LCD* and *RCD.*

Table 5: Descriptive statistics and post-hoc results for handball

	GK	LW	LB	C	RB	RW	P
C_{WID}	-	0.04 $(0.02)_{C,Bs}$	0.23 $(0.02)_{all-RB}$	0.36 $(0.03)_{all}$	0.26 $(0.02)_{all-LB}$	0.05 $(0.02)_{C,Bs}$	0.05 $(0.02)_{C,Bs}$
C_{WOD}	0.01 $(0.00)_{all-LW,P}$	0.04 $(0.02)_{C,Bs}$	0.23 $(0.02)_{all-RB}$	0.38 $(0.03)_{all}$	0.26 $(0.02)_{all-LB}$	0.05 $(0.02)_{C,Bs,GK}$	0.03 $(0.01)_{C,Bs}$
C_{WB}	-	0.03 $(0.04)_{C,Bs}$	0.23 $(0.05)_{all-RB}$	0.45 $(0.05)_{all}$	0.25 $(0.10)_{all-LB}$	0.02 $(0.02)_{C,Bs}$	0.02 $(0.02)_{C,Bs}$
C_{WC}	0.04 $(0.01)_{all}$	0.14 $(0.02)_{all-P,RW}$	0.18 $(0.01)_{all-RB}$	0.18 $(0.01)_{all-Bs}$	0.18 $(0.01)_{all-C,LB}$	0.15 $(0.01)_{all-LW}$	0.13 $(0.02)_{all-LW}$

Subscripts indicate to which tactical positions given value is statistically different for $p < .05$, e.g. *C*: given value is statistically different to the value of the center; *All:* value is statistically different to all other tactical positions; *All-"tactical position(s)"*: value is statistically different to all other tactical positions *except* the listed ones, e.g. All-C: given value is statistically different to all other values but the one of the center; *Bs* includes *LB* and *RB*.

The corresponding results for football matches under investigation can be seen in Table 4. The DM position scores the highest C_{WID} and C_{WOD} values, meaning that this position had on-average the highest number of successfully received and executed passes. Statistically significant differences can only be shown in comparison with the GK position for C_{WID} and certain attacking positions for C_{WOD} additionally. DM is also leading the C_{WB} scores followed by the RD and central defender positions. Their respective values are significantly different to the values of the other tactical positions; whereas the C_{WC} values are similar between all tactical roles apart from the GK.

The C_{WID}/C_{WOD} ratios in Figure 1 show values below 1 for defensive positions and above 1 for offensive positions, especially strikers.

For handball, C is significantly more central than all other tactical positions based on C_{WID}, C_{WOD} and especially C_{WB}. The C_{WB} values indicate that C frequently functions as the bridging unit between other tactical positions. Table 5 shows that the remaining back positions (LB and RB) have similar values for each metric and are significantly different to all other tactical positions for C_{WID}, C_{WOD} and C_{WB}. The same applies for the wing positions (LW and RW). However, their values fall into the same category with the pivot position. The GK values are neglecting low and ranked last for the considered metrics.

The C_{WID}/C_{WOD} ratios in Figure 1 reveal a high value above 1 for the point. Its share in pass reception outweighs share in pass completion.

Team Parameters

The descriptive statistics and post-hoc results for the team metrics in Table 6 show that the considered sports have significantly different values for almost all centralization measures

Table 6: Descriptive statistics and post-hoc results for team metrics

	Basketball	Football	Handball
C_{WIDC}	0.13 (0.05)$_{FB,HB}$	0.05 (0.02)$_{BB,HB}$	0.24 (0.04)$_{BB,FB}$
C_{WODC}	0.10 (0.04)$_{FB,HB}$	0.05 (0.01)$_{BB,HB}$	0.25 (0.03)$_{BB,FB}$
C_{WBC}	0.89 (0.15)$_{FB,HB}$	0.22 (0.09)$_{BB,HB}$	0.35 (0.06)$_{BB,FB}$
C_{WCC}	0.13 (0.03)$_{FB,HB}$	0.05 (0.01)$_{BB}$	0.05 (0.01)$_{BB}$

Subscripts indicate to which team sport given value is statistically different for $p < .05$, e.g. *FB*: given value is statistically different to the value in football.

calculation, we were able to follow Freeman's definition in our between each sport. As the highest values were unique in every computations. The average C_{WIDC} and C_{WODC} values are highest for handball, followed by basketball in second place. This order for first and second rank switches between these two sports for C_{WBC} and C_{WCC}. Football has the lowest average values for all team metrics employed in this study.

Visualization

Figure 2 displays the aggregated passing distribution of all matches in each sport and the corresponding MSTs next to that on the right-hand side. As edge weights were unique in each network, the resulting MSTs are unique as well (Li, Hou & Sha, 2005). The tree representing the passing network in basketball shows a typical star network topology with the PG as the central node to which all other tactical positions are connected. The topology of the handball MST has a strong resemblance with the tactical formation of the sport. The C position emerges as the centrally located node connected to the pivot and back positions who themselves are adjoined to the wings. No

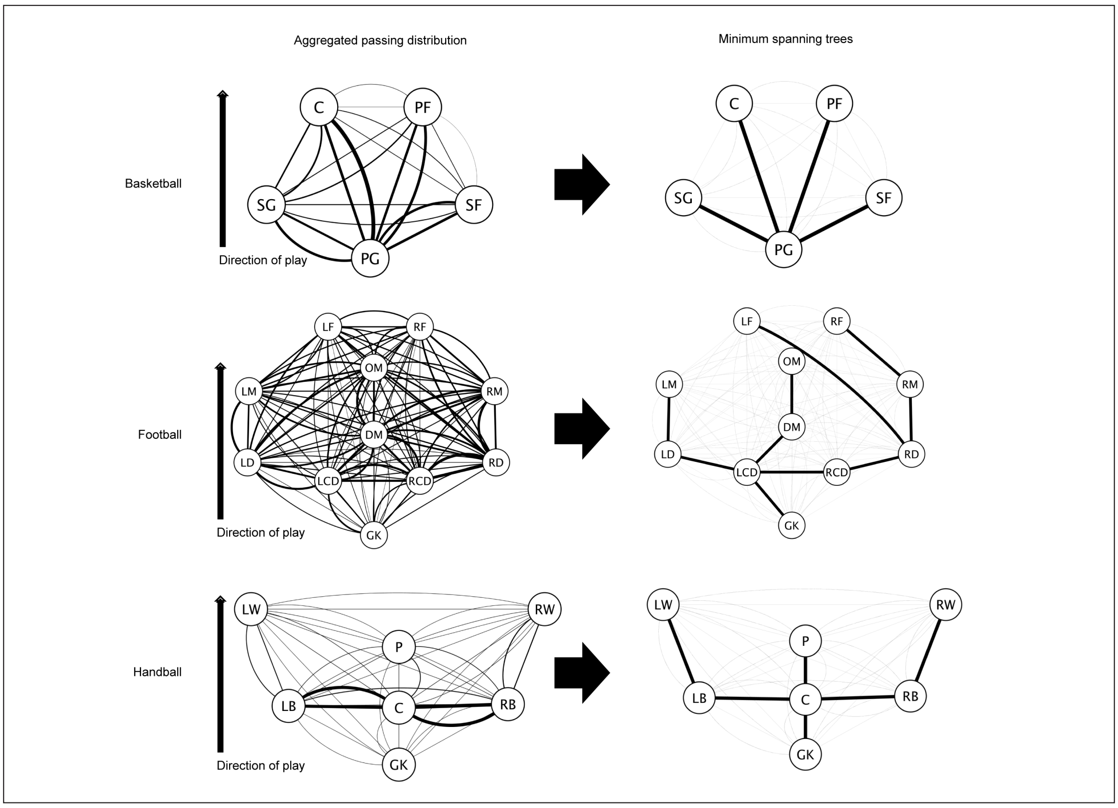

Figure 2: Visualization of aggregated passing distribution and MSTs for basketball, football and handball

distinct shape can be taken from the football MST. However, defensive positions are centrally located, and the tree displays three clusters in the longitudinal direction. Apart from the direct connection between the RD and LF, tactical positions are subdivided into left, central and right areas of the pitch and were shown as directly connected.

Discussion

The aim of this study was to characterize and compare the complex interactions visible in team sports. Network properties aid in breaking down this complexity and assessing the overall cooperation or collective organization of players and their individual contribution to a team's interaction. This is known to be vital in the analysis of team sports (Vilar et al., 2013).
This research study was conducted using passing data from several matches of major professional tournaments in basketball, football and handball. Of course, team interactions might also take other forms than passing events to express the relationship between players, e.g. the communication between the players on the field. Although there is no doubt on the importance of these forms of interaction, we assess direct passes between players as the most relevant form of interaction to characterize collective organization in team sports (Grund, 2012). The resulting analysis of our study reveals statistical differences in the pattern of play between different sports and the tactical positions therein with moderate to strong effect sizes.
The results of the individual metrics identified the DM as the most prominent player in football. He and the central defenders who act as the bridging players, as revealed by their leading C_{WB} scores, secure the ball circulation. The MST topology supports this line of argument, as these positions are centrally located within the tree, implying a strong contribution to the interaction in the sport. A centrally located player in the MST indicates a close connection or interaction with team members supporting the argument that he is a vital part in forming the collective organization of his team. There are several reasons why the RD position is also ascribed a central role to in this study according to the network metrics. First, 50% of all attacks on average were built via the right wing in comparison to 31% via the left wing. Second, the RD was among the top 3 pass executers in 10 out of 16 networks confirming the involvement of that position in building attacks via the right wing. Third, renowned players such as Philipp Lahm took on the RD position during the tournament. He alone produced 10-20 deliveries or solo runs into the attacking third per game in comparison to 2-5 for his counterpart on the LD position. This supports the dominant role of the RD and strong connection to forward positions visualized trough the connection in the MST. However, the similar C_{WC} scores suggest that all players in general are equally strongly connected with each other, directly or indirectly, implying that a quick ball circulation from any player to another is given in football, in line with previous studies (Pena & Touchette, 2010).
In basketball the central role of the PG becomes obvious looking at the C_{WB} scores. A majority of the strongest connections between positions run via the PG, identifying him as the bridging player between tactical positions in basketball. The star network topology of the MST with the PG situated in the center visualizes these findings. The dominant role of this tactical position is also in line with several previous studies (Clemente et al., 2015a; Fewell et al., 2012).
In handball, the C_{WB} results suggest a central role of the C position in facilitating the ball and structuring the interplay in that sport. The C_{WC} metric evaluates how closely a player is connected with *all* other players. The fact that the corresponding C_{WC} share is less than half as high (0.18 to 0.45) suggests that C predominantly interacts with a subset of players i.e. the back positions. The C_{WID} and C_{WOD} scores support the argument that the back positions are the dominating players here.
A deeper role division can be taken from the reported C_{WID} / C_{WOD} ratios. In football, the ratios indicate a subdivision between attacking and defensive roles. The defensive roles show higher C_{WOD} than C_{WID} values, thus ratios below 1, as they initiate plays while attacking roles rather finish them. This observation is not made in the other two sports. Solely in the case of handball, the P has a relatively high C_{WID} / C_{WOD} ratio as that player is mostly targeted to finish attacks rather than initiating them. Apart from these indications, a clear division into distinct roles is not visible in either basketball or handball. Although we analyzed matches from tournaments at the highest professional level, differences in C_{WID} and C_{WOD} values might also be ascribed to limited technical abilities to a certain extent. Whereas in basketball (13.5 turnovers against 253.7 passes for a 94.9% passing success rate on average per match for each team) and handball (10.8 turnovers against 503.4 passes for a 97.9% passing success rate) this aspect might be considered rather negligible, the passing success rate in football for the considered matches is only at 76.5%. Therefore, technical limitations might add to the high ratios of C_{WID} to C_{WOD} in football for some players.
The results of the team metrics show that general interplay is most balanced between players in football based on the distribution of all individual metrics among tactical positions. As the DM and RD have relatively high C_{WB} scores in comparison to the other tactical positions, the corresponding C_{WBC} value is slightly higher than for the other team metrics in football. This could mean, that although interplay is quite balanced, there is a tendency towards a few players having a stronger influence on the structuring of the interplay.
The interplay in basketball was demonstrated to be more unbalanced than in football. Although pass reception and execution were equally distributed between most tactical positions, the PG leads both categories significantly also resulting in higher C_{WIC} and C_{WOC} values than in the case of football. The bridging player characteristic of the PG also explains the high C_{WBC} score of 0.89. In fact, in 9 of the 16 networks in basketball the C_{WBC} score takes on the maximum value of 1. This implies that every strongest connection between any two players in these mat-

ches involved the PG confirming the dominant role of this player in facilitating the interplay.

The most unbalanced interplay between tactical positions in this study can be seen in handball according to the distribution of the direct interplay captured in the C_{WIC} and C_{WOC} scores. However, the low C_{WCC} score suggest that, similar to football, all players in handball, are quite equally strongly connected, directly or indirectly, with each other. The low direct involvement of the GK in the interplay is partly offset by the consideration of indirect connections in this metric.

The topology of the MSTs, which reduces the complexity to the most intense connections between players, offers a richer insight into certain patterns of play. For handball, the patterns in question perfectly resemble the basic order of the tactical line-up. This suggests that interplay is quite structured and predefined and therefore that the central role of the three back positions is primarily a result of their tactical position in a quite static basic order. They are crucial for the ball circulation and structure the collective organization of the team in order to score. In football, we have similar findings, however, less strong. Here a longitudinal clustering, meaning a subdivision into attacking wings, is visible. The basic order of the tactical positions appears to foster a stronger interplay of certain dyads e.g. between wing defenders and wing midfielders.

In basketball, the central role of the PG in structuring the offensive plays outweighs any other potential cluster formation of tactical positions, resulting in the star network topology of the MST. According to Bonanno, Caldarelli, Lillo and Mantegna (2003) this kind of topology is an argument for a clear hierarchical structure, i.e. that the PG has a strong impact on structuring the interplay of his team. Teammates continuously bring the PG into possession to initiate and structure plays (Bourbousson, Poizat, Saury & Seve, 2010).

The main limitation seen in this research study was related primarily to the sample size of the data utilized. Moreover, matches from only one major tournament are considered in each sport. In order to generalize the results for each sport, a larger sample across different occasions would be needed. Besides, definitions of tactical positions in football are approximations in some instances by combining data on tactical lineups and positional data provided by FIFA (www.fifa.com/worldcup/archive/brazil2014). There is an overall consensus on the definition of tactical roles in previous studies focusing on basketball and especially handball induced by its quite static formation (Cardinale, Whiteley, Hosny, & Popovic, 2017; Fewell et al., 2012; Karcher & Buchheit, 2014). However, in football, we acknowledge that tactical roles are a more complex factor. Here, we believe that temporarily occupying different areas on the pitch and fulfilling different tasks, i.e. a striker who takes on defending tasks, can be acknowledged as part of the role repertoire of players in football. Eventually, this is why we are faced with such complex webs of interaction in which different tactical positions interact with each other and that network analysis is able to capture for the purpose of our study.

Moreover, it is important to make two remarks regarding the application of weighted closeness in this study. First, one could argue that the nearly completeness of the present networks in this study, in which almost all players are directly connected with each other, mostly account for the similar C_{WC} scores in football. However, in basketball, for example, we find statistical differences especially with regard to the PG while having complete networks in every analyzed match exclusively. We claim that in weighted networks, in comparison to unweighted networks, strong indirect connections might dominate weak direct connections and thus weaken the influence of the level of completeness in a network to a certain degree.

Second, only 13 of the 16 analyzed networks could be considered in the one-way ANOVA of the C_{WC} scores in handball, as the GK was not involved in any interplay in some matches. However, as the metric analyzes the connection with *all* players in the network and cannot consider disconnected components by definition, we had to drop three networks (Opsahl et al., 2010). This stresses the low involvement of the GK in building attacks in handball.

Nevertheless, this study contributes to the understanding of the nature of team sports and the respective involvement of the different tactical positions within each sport. This identifies SNA as a powerful tool not only to break down the performance of a single sport but also to allow a profound comparison between the styles of interaction in team sports.

Conclusion

The aim of this study was to characterize the nature of team sports and the role of their respective tactical positions.

By applying methods from social network analysis it was possible to break down the complexity of a handful of popular sports, by quantifying and intuitively visualizing roles of players and overall team interaction. Thus, this is the first study that compares the network patterns of different team sports. Moreover, MSTs are applied for the first time in a team sports context which in particular turn out to be effective in breaking down the complexity of almost complete networks.

Ultimately, the analysis revealed significant findings, on the prominent tactical positions for building attacks in the three sports discussed: in basketball, this dominant tactical position tended to be the PG, in football the DM and C in handball. The general pattern of play appears to be significantly more unbalanced in handball than in basketball and football. As a final takeaway, the study indicated strong findings that the level of fixedness in the basic order of the tactical positions in the sports influences the prominence levels of players.

We chose three popular invasion games in this study to offer a first comparison between the network properties of team sports. However, as we assess the outlook of this method as fruitful, more team sports should be incorporated in future studies to further examine and characterize the different dynamic systems present in team sports. Moreover, individual modifications of traditional network metrics may lead to an even more accurate quantification of performance in each sport.

Funding

The authors have no funding or support to report.

Competing Interests

The authors have declared that no competing interests exist.

Data Availability Statement

All relevant data are within the paper.

References

Araújo, D., & Davids, K. (2016). Team synergies in sport: Theory and measures. *Frontiers in Psychology*, *7:1449*. doi: 10.3389/fpsyg.2016.01449

Bonanno, G., Caldarelli, G., Lillo, F., & Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, *68* (4), 046130.

Bourbousson, J., Poizat, G., Saury, J., & Seve, C. (2010). Team coordination in basketball: Description of the cognitive connections among teammates. *Journal of Applied Sport Psychology*, *22* (2), 150–166.

Cardinale, M., Whiteley, R., Hosny, A. A., & Popovic, N. (2017). Activity profiles and positional differences of handball players during the world championships in Qatar 2015. *International Journal of Sports Physiology and Performance*, *12*, 908–915.

Clemente, F. M., Martins, F. M. L., Wong, P. D., Kalamaras, D., & Mendes, R. S. (2015a). Midfielder as the prominent participant in the building attack: A network analysis of national teams in FIFA World Cup 2014. *International Journal of Performance Analysis in Sport*, *15*, 704–722.

Clemente, F. M., Martins, F. M. L., Kalamaras, D., & Mendes, R. S. (2015b). Network analysis in basketball: Inspecting the prominent players using centrality metrics. *Journal of Physical Education and Sport*, *15*(2), 212.

Clemente, F. M., & Martins, F. M. L. (2017). Network structure of UEFA Champions League teams: Association with classical notational variables and variance between different levels of success. *International Journal of Computer Science in Sport*, *16*(1), 39–50.

Courneya, K. S., & Carron, A. V. (1992). The home advantage in sport competitions: A literature review. *Journal of Sport and Exercise Psychology*, *14*(1), 13–27.

Duch, J., Waitzman, J. S., & Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PloS one*, *5*(6), e10937.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532.

Fewell, J. H., Armbruster, D., Ingraham, J., Petersen, A., & Waters, J. S. (2012). Basketball teams as strategic networks. *PloS ONE*, *7*(11), e47445.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Gama, J., Passos, P., Davids, K., Relvas, H., Ribeiro, J., Vaz, V., & Dias, G. (2014). Network analysis and intra-team activity in attacking phases of professional football. *International Journal of Performance Analysis in Sport*, *14*(3), 692–708.

Glazier, P. S., & Davids, K. (2009). Constraints on the complete optimization of human motion. *Sports Medicine*, *39*(1), 15-28.

Gower, J. C., & Ross, G. J. (1969). Minimum spanning trees and single linkage cluster analysis. *Applied Statistics*, *18*, 54–64.

Grund, T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, *34*, 682–690.

Gwet, K. (2001). *Handbook of inter-rater reliability: How to estimate the level of agreement between two or multiple raters*. Gaithersburg, MD: STATAXIS.

Karcher, C., & Buchheit, M. (2014). On-court demands of elite handball, with special reference to playing positions. *Sports Medicine*, *44*, 797–814.

Lappas, T., Liu, K., & Terzi, E. (2009, June). Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 467–476). ACM.

Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the Association for Information Science and Technology*, *58*, 1303–1319.

Li, N., Hou, J. C., & Sha, L. (2005). Design and analysis of an MST-based topology control algorithm. *IEEE Transactions on Wireless Communications*, *4*, 1195-1206.

Li, C. T., & Shan, M. K. (2010, August). Team Formation for Generalized Tasks in Expertise Social Networks. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing* (pp. 9–16). IEEE Computer Society.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, *11*(1), 193–197.

Newman, M. E. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical review E*, *64*(1), 016132.

O'Donoghue, P. (2009). *Research methods for sports performance analysis*. London: Routledge.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251.

Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., & Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport*, *14*(2), 170–176.

Passos, P., Araújo, D., & Volossovitch, A. (2016). *Performance analysis in team sports*. London: Routledge.

Pena, J. L., & Touchette, H. (2012). A network theory analysis of football strategies. *arXiv preprint arXiv:1206.6904*.

Vilar, L., Araújo, D., Davids, K., & Bar-Yam, Y. (2013). Science of winning soccer: Emergent pattern-forming dynamics in association football. *Journal of Systems Science and Complexity*, *26*(1), 73–84.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: University Press.

A.3 Study 2 - Permission for the inclusion in dissertation

Wednesday, July 17, 2019 at 10:46:06 AM Central European Summer Time

Betreff: RE: review process - JHK
Datum: Mittwoch, 17. Juli 2019 um 10:43:42 Mitteleuropäische Sommerzeit
Von: Journal of Human Kinetics
An: Korte, Florian

Dear Florian Korte,
We do confirm that it is allowed to include the accepted article "Passing network analysis of positional attack formations in handball" within your dissertation, yet please note that providing a reference would be necessary.
With kind regards,
Aleksandra Mostowik

Aleksandra Mostowik
Editorial Office of Journal of Human Kinetics
Academy of Physical Education in Katowice
e-mail: jhk@awf.katowice.pl

From: Korte, Florian [mailto:korte@cdtm.de]
Sent: Friday, July 12, 2019 1:46 PM
To: Journal of Human Kinetics
Subject: Re: review process - JHK

Dear Editorial Team,

following the Open-Access Policy, I would need a quick confirmation that I can include the accepted article "Passing network analysis of positional attack formations in handball" within my dissertation. If you could please provide me with a short confirmation that this is ok, that would be very helpful. Thank you very much!

Best regards,
Florian Korte

A.4 Study 2 - Original article

Passing network analysis of positional attack formations in handball

Abstract

The aim of this study was to characterize handball from a social network analysis perspective by analyzing 22 professional matches from the European Men's Handball Championship 2018. Social network analysis has proven successful in the study of sports dynamics to investigate the interaction patterns of sport teams and the individual involvement of players. In handball, passing is crucial to establish an optimal position for throwing on goal. Moreover, different tactical formations are played during a game, often induced by two-minute suspensions or the addition of an offensive player replacing the goalkeeper as allowed by the International Handball Federation since 2016. Therefore, studying the interaction patterns of handball teams considering the different playing positions under various attack formations contributes to the tactical understanding of the sport. Degree and flow centrality as well as density and centralization values are computed. As a result, a quantification of the contribution of individual players to the overall organization is achieved alongside the general balance in interplay. We identify the backcourt as the key players to structure interplay across tactical formations. While attack units without a goalkeeper are played longer, they are either more intensively structured around back positions (7vs.6) or spread out (5+1vs.6). We also find significant differences in the involvement of wing players across formations. The additional pivot in the 7vs.6 formation is mostly used to create space for back players and is less involved in interplay. Social network analysis turned out as a suitable method to govern and quantify team dynamics in handball.

Keywords

social network analysis, temporal networks, centrality measures, performance analysis, tactical analysis, team sports

Introduction

Matches in team sports are complex dynamic systems caused by the frequent interaction between players (Glazier and Davids, 2009). Teams work together collectively to achieve the common goal of winning (Lusher et al., 2010). In fact, the synchronized action of players in a team is regarded as a crucial part of the key factors to successful performance (Grund, 2012). Here, passing, which is a common performance variable in notational analysis of team sports, is the foundation for the collective action of players in a team (Passos et al., 2016).

In handball, ball circulation is crucial to establish an optimal position for throwing on goal (Wagner et al., 2014). However, varying environmental constraints, such as the configuration of the opposing line-ups, require different interaction patterns in order to succeed (Araújo and Davids, 2016). There is a set of different tactical formations that are played during a handball game, often induced by two-minute suspensions or the addition of an offensive player replacing the goalkeeper as allowed by the International Handball Federation (IHF) since 2016. Therefore, studying the interaction patterns of handball teams considering the different playing positions under various attacks formations contributes to the understanding of the sport and its actual development.

Social network analysis (SNA) has proven successful in the study of ball passing dynamics by breaking down the complexity within the web of interactions between players (Passos et al., 2011). As a match analysis tool, SNA is able to quantify the contribution of individual players to the general interplay as well as detecting patterns in the passing structure of teams (Clemente and Martins, 2017).

On a mirco-level, focusing on individual performance, professional matches in football have been analyzed predominantly. As midfielders are responsible for building attacks, they are identified as the most prominent players in the majority of studies (Clemente et al., 2015a; Pena and Touchette, 2012). Clemente and Martins (2017) also consider different tactical formations in their computation of network metrics in professional football. On a macro-level, focusing on general team performance, several studies suggest a strong correlation between successful team performance and frequent but also balanced interplay between players (Clemente et al., 2015b; Duch et al., 2010; Grund, 2012).

In summary, most of the studies in SNA are conducted in football. Most studies in handball, however, rather discuss physical and technical attributes of the sport (Karcher and Buchheit, 2014; Michalsik and Aagaard, 2015; Póvoas et al. 2012). Tactical components, especially in terms of interplay, have not been studied extensively yet. Korte and Lames (2018) offer a first insight into the interplay in handball on an aggregate match-level. They identify the backcourt players as most central in terms of structuring interplay but do not account for different tactical formations. The newly introduced option to replace the goalkeeper in attacking phases alongside the frequent occurrence of temporary suspensions has enriched the sport with an extensive set of attacking formations or constellations and thus varying constraints for attacking teams. According to Gruic et al. (2006) the resulting tactical setup during attack phases influences the interplay of attacking teams. In particular, they found that backcourt players adapt their passing behavior according to changes in the tactical setup. However, their analysis is rather qualitative and tactical formations under the new IHF rule are not considered. Hence, analyzing the collective organization of teams and their passing patterns during different types of attack phases is very important to better understand the sport of handball.

Therefore, the aim of this study was to characterize interplay by focusing on positional attacks across different tactical formations. As we also differentiate between playing positions, the focus did not only lie on the general structure of interplay but also on the individual contribution of players within a team. On a micro-level, we calculated weighted in-/out-degree and flow centrality to assess the overall involvement of playing positions in attacks across a match and their contribution to structuring plays within attack units. On a macro-level, density and weighted degree centralization was computed to better understand the level of cohesion between players and balanceness of interplay.

To our knowledge, this is the first study that attempts a tactical analysis of interplay in handball by differentiating between prevalent tactical formations as well as playing positions exploiting metrics of SNA. Moreover, it pioneers the breakdown of the assessment on attack unit level to take the temporal component of handball into account.

Methods

Samples

A total of 22 matches of the 2018 EHF European Men's Handball Championship were analyzed in this study including all encounters from the main round, two semi-finals, the third-place match and final. A total of 3,100 directed adjacency matrices, one for each attack unit, capture an aggregated amount of 17,420 passes between players in our analysis.

Procedure

Conducting SNA requires passing networks constructed from a set of nodes and edges. The nodes represent players whereas the edge weights stand for the number of passes between them. Following Ramos et al. (2018), we conduct our analysis on attack basis instead of aggregate match level to consider the temporal character of handball. That means, instead of aggregating the passing data of a team throughout a whole match before running analysis, we evaluate each attack separately. That way, we can track and analyze actual sequences of interplay instead of average connections across a series of attacks.

When focusing on attacks, literature differentiates between counter-attacks and positional attacks in handball (Karcher and Buchheit, 2014). We only focus on the latter, meaning organized positional attacks with offensive as well as defensive players having taken their respective playing position (Yamada et al., 2014). This is because we want to focus on the structured and controlled interaction to overcome defensive lines which make up 87% (1,993 in total) of all attacks in our study. As ball possession of attacking teams is often subdivided into multiple sub-attacks or offensive attempts, caused by a referee decision, throw-in or repossession of a deflected ball (Pfeiffer and Perl, 2006), we define these as our smallest units of attack (3,100 in total) to most accurately represent the concept of interplay (Ramos et al., 2018).

To characterize the different types of sub-attacks we differentiate between four common tactical formations, namely 6vs.6, 6vs.5, 5+1vs.6 and 7vs.6. Whereas the first number describes the number of offensive players, the second number states the number of defensive players within the sub-attack, accordingly. 6vs.6 can be seen as the most common *base formation*, 6vs.5 implies a two-minute penalty in the defending team, 5+1vs.6 reflects a two-minute suspension in the attacking team which is compensated by replacing the own goalkeeper with an additional attacking player. The tactical formation 7vs.6, on the other hand, implies a goalkeeper replacement by the attacking team, as described above, without having suffered a temporary suspension. In our study, most of the sub-attacks are played in a 6vs.6 formation (74.5%), 10.3% in 5+1vs.6, 7.7% in 6vs.5 and 3.9% in a 7vs.6 formation. The remainder consists of other infrequently played formations such as 5vs.5 or 7vs.5. However, we focus on the four most frequent tactical formations that make up 96.4% of all attack units.

To better understand attack formations and be able to characterize handball as such, we track playing positions and not players (Póvoas et al., 2012). In handball, we find a clear differentiation between tactical roles (Cardinale et al., 2016). Therefore, we codify the following playing positions: i) left wing (LW); ii) left back (LB); iii) center (C); iv) right back (RB); v) right wing (RW); vi) pivot (P); and vii) an additional pivot in 7vs.6 (P7). As the goalkeeper is not involved in positional attacks, we drop this playing position from analysis.

To overcome the issue of frequent substitutions, especially in the backcourt, we reassign playing positions (Michalsik and Aagaard, 2015). The tracking and codification process is done by researchers with more than 15 years of experience in handball. It was executed through video analysis applying the software Dartfish®.

To ensure the reliability of the data, we compute Cohen´s kappa and Gwet's AC1 inter-rater statistic in a two-stage process (Gwet, 2001). Using Gwet's statistic, we first analyze the agreement on the occurrence of passes. In a second step, Cohen's Kappa tests the agreement on pass executer and receiver. Moreover, the agreement on tactical formation is tested. 15% of the overall data was assessed for reliability purposes. The Kappa (Gwet) values were above 0.95 (0.80) respectively for passing and 0.83 for the agreement on tactical formations, meeting the requirements for observer agreement (Robinson and O'Donoghue, 2007).

Network Metrics

The software Matlab® was used to carry out the analysis and the visualization of networks was enabled through Cytoscape®. A set of individual and team centrality metrics were computed. They allow a quantification of the involvement of playing positions in executing and structuring interplay as well as the overall distribution and layout of passing within an attacking team. We consider weighted directed graphs to consider both passing directions between any set of two attacking players. On a micro-level, weighted in-/out-degree as well as flow centrality are computed. On a macro-level, density and weighted degree centralization are calculated to assess the general structure of interplay in different formations.

Weighted In-Degree

Weighted in-degree, also referred to as prestige in SNA, is the sum of all incoming weighted edges of a particular node. Thus, in a handball context, it captures the number of received passes by a player during an attack unit. Let n_i be a node of weighted directed graph G with n nodes. Then, weighted in-degree index, $C_{WID}(n_i)$, for player i is calculated as

$$C_{WID}(n_i) = \sum_{\substack{j=1 \\ i \neq j}}^{n} a_{ji} \tag{1}$$

where a_{ji} corresponds to the frequency of passes from player j to i. The metric is often taken as a first indicator for the prominence of a player. A player that is being targeted frequently during an attack is mostly likely trusted by fellow players to structure the team's attacking plays (Clemente et al., 2015a, Korte and Lames, 2018).

Weighted Out-Degree

Weighted out-degree, also referred to as centrality, takes the sum of all outgoing weighted edge values of a certain node. It therefore represents the number of executed passes by a player during an attack unit. Let n_i be a node of weighted directed graph G with n nodes. Then, weighted out-degree index, $C_{WOD}(n_i)$, for player i is calculated as

$$C_{WOD}(n_i) = \sum_{\substack{j=1 \\ i \neq j}}^{n} a_{ij} \tag{2}$$

where a_{ij} corresponds to the frequency of passes from player i to j. In recent studies, this metric was often used to describe players with a high contribution to the overall ball circulation (Clemente et al., 2015a).

Both degree metrics are computed on an absolute as well as relative level. We obtain relative values as a share of the aggregated degree levels across all playing positions. Moreover, we also carry out an analysis of a subset which only includes attack units of at least three passes for these two metrics to provide a richer insight into passing patterns in handball by focusing on longer attacking plays.

Flow Centrality

Flow centrality is calculated as the fraction of passing sequences (or attack units) that a particular playing position is involved in relative to all plays of its team within a match (Fewell et al. 2012). In contrast to the weighted degree centrality metrics above, flow centrality does not assess the average involvement of a particular player within attack units, but the overall prevalence in attack units across the entire match. This enters a new aspect to the assessment of interplay. By only looking at weighted degree, the intermediary role of a player, who is highly involved in the passing of only a small set of attack units across a match, might be overestimated. In contrast, flow centrality focuses on the share of attacks that a particular player is at least once involved in. As it offers a holistic evaluation of the involvement across an entire match, it is increasingly used to assess the intermediary role of individual players (Duch et al., 2010). Flow centrality index, $C_{FC}(n_i)$, for player i is calculated as

$$C_{FC}(n_i) = \frac{\sum_{k=1}^{m} s_k(n_i)}{s_m} \tag{3}$$

where s_m denotes the total number of m attack units in a match and $s_k(n_i)$ denotes the k-th attack unit in which n_i is involved at least once. By construction, all flow centrality values are bounded between 0 (player n_i is not involved in any attack unit of its team in the match) and 1 (player n_i is involved in all attack units of its team in a match). In constrast to the concept of betweenness, which focuses on paths, it rather considers walks. Paths are based on the strongest connections in terms of pass frequency between any set of two players. However, it does not necessarily describe an actual passing sequence. In contrast, walks consider direct interplay during attack phases (Borgatti, 2005). Thus, flow centrality is seen as a more appropriate metric to describe intermediary players (Ramos et al., 2018). In addition, we examined flow centrality restricted to interactions in the final three passes before a shot-on-goal situation to study network properties in the crucial phase of an attack unit, following Fewell et al. (2012). The index for this specific metric is defined as $C_{FC3}(n_i)$ for player i.

Density

Density is the number of actual connections between attacking players as a share of the potential connections. The latter is a connection (or technically: edge) that could potentially exist between any sets of two attacking players. Thus, this metric provides a quantification of the general level of cohesion across a team within an attack unit. For the computation, we assess the direction of the pass as irrelevant as the focus purely lies on the occurrence of a connection. For a weighted digraph G with n nodes, density index, C_D, is calculated as

$$C_D = \frac{2 * \sum_{\substack{j=1 \\ i \neq j}}^{n} c_{ij}}{(n-1) * n} \tag{4}$$

where c_{ij} is an indicator function that takes the value 1 if there is at least one pass from player i to j or vice versa. Otherwise, it takes the value 0. The metric is adjusted by the total number of potential connections between n nodes.

Weighted Degree Centralization

Weighted degree centralization takes the sum of all deviations from the weighted degree values of all nodes to the highest value in the network adjusted by the number of players and passing intensity (Freeman, 1978, Opsahl et al., 2010). The weighted degree value of a node is simply the sum of its weighted in-/out-degree values. In a sports context, the metric provides an indication to what level the cohesion is concentrated around certain players of the attacking team. For a weighted graph G with n nodes, weighted degree centralization index, C_{WDC}, is calculated as

$$C_{WDC} = \frac{\sum_{i}^{n} C_{WD}^{*} - C_{WD}(n_i)}{(n-1) * C_{WD}} \tag{5}$$

where C_{WD}^{*} is the highest weighted degree value of a playing position in its team, $C_{WD}(n_i)$ the weighted degree value of playing position i and C_{WD} the aggregated weighted degree values of all playing positions, which can also be referred to as passing intensity (Grund, 2012). The adjustment according to the number of attacking players allows a comparison between tactical formations.

Statistical Procedures

For individual metrics, two-way ANOVA are carried out for each dependent variable, degree and flow centrality. Tactical formation and playing position are the independent factors of our analysis. We conduct multiple one-way ANOVA to analyze the variance within each factor and Tukey HSD post-hoc tests for pairwise comparisons between tactical formations and playing positions, respectively. For team metrics, C_D and C_{WDC}, multiple one-way ANOVA are executed to test for statistical differences between tactical formations. Our statistical analysis is conducted with Matlab at a 5% significance level. Following Ferguson (2009) and Clemente and Martins (2017), η^2 is reported to interpret the effect size according to the following criteria: no effect ($\eta^2 < 0.04$); small effect ($0.04 \leq \eta^2 < 0.25$); moderate effect ($0.25 \leq \eta^2 < 0.64$); strong effect ($\eta^2 \geq 0.64$).

Network Visualization

A visualization of the results is provided by a depiction of common network plots with nodes and edges representing playing positions and passing frequency, respectively. The 5+1vs.6, 7vs.6 and 6vs.5 formations are visualized as the relative difference values compared to 6vs.6, both positive and negative.

Results

Individual Parameters

The results of the two-way ANOVA reveal significant differences in the independent variable of playing position on C_{FC} ($p < .001\,;\eta^2 = 0.744$), C_{FC3} ($p < .001\,;\eta^2 = 0.625$), C_{WID} ($p < .001;\eta^2 = 0.163$) and C_{WOD} ($p < .001;\eta^2 = 0.197$). Moreover, significant differences are found with regard to tactical formation on C_{FC} ($p < .001\,;\eta^2 = 0.023$), C_{WID} ($p < .001;\eta^2 = 0.007$) and C_{WOD} ($p < .001;\eta^2 = 0.008$). No statistical differences are found for the independent variable of tactical formation with regards to C_{FC3} ($p = 0.121\,;\eta^2 = 0.007$). There are also statistically significant interactions between tactical formation and playing position on C_{FC} ($p < .001\,;\eta^2 = 0.056$), C_{FC3} ($p = 0.003;\eta^2 = 0.040$), C_{WID} ($p < .001;\eta^2 = 0.008$) and C_{WOD} ($p < .001;\eta^2 = 0.006$) including the filtered subset focusing on attacks of at least three passes. That means formation changes affects playing position involvement, measured by our individual metrics, differently.

Table 1 insert here.

The results of the one-way ANOVA demonstrate significant effects between centrality levels of playing positions for each tactical formation with respect to all individual centrality measures. Table 1 shows that the highest average values were found for the center position, C, followed by both back positions (LB and RB) with respect to all relevant centrality measures. Wings (LW and RW) and pivot(s) (P and P7 for 7vs.6) scored lowest for each tactical formation. Focusing on flow centrality, C is involved in at least 94% of all attacking interplays for each tactical formation and at least 92% when focusing solely on the final three passes before a shot on goal. No significant differences were found between the back positions here. The pivot, P, is significantly more involved in attack units than the wing positions, apart from the 6vs.5 formation in which the LW and RW are more prevalent. Between those two playing positions we only find significant differences within the 6vs.6 formation. The additional pivot, P7, is only part of the interplay in about 2% of all attack units taking place in a 7vs.6 formation.

The results of the multiple one-way ANOVA of C_{WID}, C_{WOD} , C_{FC} and C_{FC3} per playing position for each tactical formation can also be taken from Table 1. For C and the other two back positions, LB and RB, absolute C_{WID} and C_{WOD} values are significantly higher in a 5+1vs.6 and 7vs.6 formation than in 6vs.6 and 6vs.5, respectively. Figure 1 shows that for attacking plays with more than three passes, there are no significant differences between 5+1vs.6, 7vs.6 and 6vs.6 for the three back positions anymore. C_{WID} and C_{WOD} values of wing players are lowest in the 7vs.6 formation, independent of the length of the attacking plays, though only partly significantly. They score highest in the 6vs.5 (RW) and 5+1vs.6 formation (LW). In general, 5.5 passes are played in a 6vs.6 formation per sub-attack, 6.9 passes (+25.5%) in a 5+1vs.6 setup, 6.5 passes (+18.2%) in the 7vs.6 constellation while there are only 4.6 passes (-16.4%) on average in a 6vs.5 formation.

Figure 1 insert here.

Focusing on the relative shares in C_{WID} and C_{WOD} values across tactical formations, we only find few significant differences, as visualized in Figure 1. However, the center position has a significantly lower share in received passes in attack units played in a 6vs.5 formation, whereas the wing players show significantly higher values during these attack phases in comparison to the other formations.

Our analysis also shows significant differences in the overall attack involvement (C_{FC} and C_{FC3}) per playing position for each tactical formation. However, the ranking across tactical formations in terms of the individual metric values is mixed for playing positions. When focusing on the final three passes of an attack, we only find significantly higher flow centrality values for the LB in a 7vs.6 formation against 6vs.6 and 5+1vs.6 as well as the RW in a 6vs.5 formation against 6vs6. and 5+1vs.6.

Team Parameters

We find significant differences between tactical formations for C_D $(p < .001; \eta^2 = 0.024)$ and C_{WDC} $(p < .001; \eta^2 = 0.013)$. On average, the density values (0.17) are significantly higher and centralization values (0.33) significantly lower in a 5+1vs.6 formation than in the all others, though with nearly no effect size. The average density value of 0.17 implies that 17% of possible connections between the attacking players are utilized for interplay which amounts to 2.4 of the 15 potential connections on average. The weighted degree centralization value is highest within the 7vs.6 formation, though not significantly different from the other formations. Table 2 visualizes the results of our analysis.

Table 2 insert here.

Network Visualization

The aggregated passing distribution between playing positions in Figure 2 confirms the relatively lower share in passing of the C and higher share for wing positions in the 6vs.5 formation. Moreover, it visualizes the increased prevalence of LB in attacking plays and low involvement of wing positions in the 7vs.6 formation compared to 6vs.6.

Figure 2 insert here.

Discussion

The study reveals statistical significance with respect to differences of centrality measures, at both micro and macro level, between tactical formations and playing positions. Effect sizes found were small to moderate.

Across the four most prevalent tactical formations in handball, the overall involvement of playing positions in attack units per match and their average passing involvement per attack unit varies differently. Our analysis shows that the effect is mostly moderated by differences between playing positions within each formation and less by substantially changing centrality levels of individual playing positions across different formations. Here, we also detect significant differences. However, effect sizes were small to negligible.

We find that interplay is dominated by and structured around the three back positions C, LB and RB across all formations. This is in line with Srhoj et al. (2001) who finds that this is partly induced by the favorable position on court which is also prevalent in all tactical formations. The dominance is demonstrated in the explicitly high flow centrality values indicating an almost persistent involvement in each attack unit, while wing and pivot players are only involved in every third or fourth positional attack unit. One explanation for these findings is that the attack efficiency in handball was found to decrease with an increasing duration of positional attacks (Rogulj et al., 2011). Towards the beginning of an attack the opposing team might struggle to form an effective defense which offers back players an easier scoring opportunity. According to the authors players in back positions therefore attempt to finalize attacks as early as possible and, thus, often without the inclusion of wing or pivot players. The high C_{WID} and C_{WOD} values underline that the backcourt is not only more prevalent in attacks during the match but also structures them within. The average numbers of pass execution and reception are highest for C, who can be seen as the key player in structuring plays, followed by the back players. Wing and pivot players have similar passing numbers on average but are significantly less involved in structuring interplay. This is in line with Foretic et al. (2013) who, in their study on situational efficiency in men's top-level handball, ascribe back players the task of organizing the game with the aim of creating a favorable position for attack finalization. The resulting three hierarchy layers of centrality also reveal a symmetric level of involvement between left- and right-sided players, especially in regard to LB and RB.

Although a mutual hierarchy is visible among playing positions in terms of interplay involvement across formations, a closer look at the results of the passing statistics and centrality metrics also shows differences in passing behavior and general interplay between tactical formations. First, they differ in their average number of passes per positional attack. Attack units with no goalkeeper such as 7vs.6 and 5+1vs.6 are played significantly longer on average (+20%) than 6vs.6 and 6vs.5. One explanation for that finding could be that teams in a 5+1vs.6 formation intend to lapse time while playing in minority. As we find a similar result for the 7vs.6 formation, the missing goalkeeper could also be a factor. Teams might avoid sudden shot attempts as they fear an almost certain turnover goal and hence decide to rather pass on the ball. For the 6vs.5 formation, in contrast, it is most likely that attacking teams either want to efficiently exploit their majority play or are simply able to quicker find the necessary gaps in the decimated defense, both resulting in shorter passing sequences on average.

Combining the passing statistics with the results from the team metrics offers a richer insight into understanding the style of interplay. The density values are significantly higher in the 5+1vs.6 than other formations meaning that more potential connections between players are exploited in this formation. However, the magnitude is quite small and does not even add up to a complete additional connection on average in comparison to the other formations. The centralization values are quite balanced and differences are low in magnitude and effect size implying that the concentration of interplay around certain focal points is balanced between formations. However, it is important to point out that by construction of the centralization metric the highest average value across formations, which is documented for the 7vs.6 formation, might underestimate the true level of concentration around crucial positions. The adjustment due to the higher number of attacking players naturally decreases its centralization value especially as passing involvement of the additional pivot, P7, is neglectingly low. This is a first hint, that, although interplay takes on average longer in the 7vs.6 formation, it is in fact more concentrated around the back positions in contrast to the other formations.

To better understand the impact of playing positions on interplay, it is crucial to look at the differences in individual centrality metrics per playing position for each formation. The number of executed and received passes of the three back positions are significantly higher in 5+1vs.6 and 7vs.6 than in the other two formations. As the relative degree values of the backs and C remain quite stable across formations, it is evident that the longer average passing is evenly structured around these three particular playing positions. What turns out to be different between interplay in the two formations that replace their goalkeepers is their different levels of inclusion of wing and pivot positions. Wing players in the 5+1vs.6 formation show significantly higher passing values than in 7vs.6. This supports the argument that longer passing sequences and the higher level of cohesion in 5+1vs.6 is also used to spread interplay to wings. In contrast, wing positions face the lowest values in a 7vs.6. formation. Whereas the passing involvement of the (standard) pivot position is even between formations, the additional pivot from 7vs.6 is nearly never targeted for interplay. Instead, it appears that its role is that of a blocker to provide better shooting opportunities for the back positions. This assumption is also supported by the significantly higher involvement of the LB, which is often referred to as the key shooting position, in 7vs.6 attack units (Karcher and Buchheit, 2014). This is especially true for the final three passes before a shot on goal.

Turning to the basic 6vs.6 formation, one would expect that the interplay is quite similar to the 5+1vs.6 formation given that the same numbers of attacking and defending players face each other. Indeed, the respective involvement of players in attacks and the relative share of involvement in interplay per attack unit is similar or nearly exact. The main difference lies in the significantly higher average number of executed and received passes of three backcourt players in 5+1vs.6. As the average increase in passing is mostly spread across three positions, and also wing and pivot players are stronger involved, the difference is not detectable in the relative shares of involvement and centralization values. Thus, the general interplay structure, especially with respect to balanceness, is similar. However, the significant differences in interplay involvement completely neutralize for backcourt players when filtering for attack units with at least three passes. It suggests that the lower average number of passes in the 6vs.6 formation is mostly due to its higher share of positional attack units with less than three passes. In fact, the share amounts to 30.7% in 6vs.6 while it is 16.6% in 5+1vs.6. Once longer offensive plays are initiated, there are no significant differences in general structure and interplay.

The 6vs.5 formation stands out as the most different from the others in terms of interplay. In contrast to the majority interplay in 7vs.6, plays are structured shorter in 6vs.5, which can be seen by the low average passing number and 26.2% share of attacks that take less than three passes. It appears, that exploiting gaps in the decimated defense is easier. Moreover, wing positions are involved quite frequently as supported by two arguments. First, the involvement ratio in the last three passes before a shot on goal is highest in the 6vs.5 formation. Second, these playing positions show increased C_{WID} values, implying that wing positions are passed to more frequently, most likely to spread interplay and create open space on the wings as an alternative to breaking through in the backcourt.

The main limitation seen in this research study was related to the unbalanced prevalence of different attack formations in the European Championship with the 6vs.6 formation adding up to most of the attack units. A bigger sample size might increase the prevalence of other attack formations. Second, the different number of attacking players (P7 only present in 7vs.6 formation) had a slight effect on the computation of the team metrics which naturally increased the complexity of our comparison. This should be noted in future research on other team sports such as field hockey or water polo, which also have temporary suspensions that influence the number of active players on the pitch. Moreover, this study focuses on the passing interaction leading towards a favorable shooting position. As it does not include the attack outcome itself, it does not break down the actual shooting performance. In general, it is important to stress that neither the situational efficiency of playing positions is assessed nor a differentiation between specific attack models provided. Similarly, defense formations during positional attacks, which could potentially impact the ball passing dynamics, are not considered.

Conclusion

The aim of this study was to characterize the nature of interplay in handball through analyzing passing sequences of positional attacks in the most prevalent tactical formations. By applying centrality metrics from social network analysis, we can quantify the involvement of playing positions and assess the playing style within different formations. Thus, this is the first study that offers a profound analysis of interplay in handball especially under the consideration of the new constraint of goalkeeper replacement in attacking plays. Moreover, our analysis, for the first time in handball, breaks down the complexity of interplay to separate attack units and thus considers actual passing sequences instead of average connections.

The main findings of this study were the significant differences in the attack involvement between playing positions across the most prevalent tactical formations. Attacking plays are predominantly structured by the C and back positions, regardless of the tactical lineup. Average passing sequences are longest in attack formations without a goalkeeper and shortest in the 6vs.5 majority formation. Whereas longer plays in 7vs.6 are mostly structured around back positions, interplay in 5+1vs.6 includes wing positions more frequently. The highest level of inclusion of wing players is found within the 6vs.5 formation, most likely to exploits gaps in the decimated defense.

Future studies should consider variations in the tactical behavior of defensive formations to more accurately account for the dynamic processes taking place between opposing teams in handball.

Ultimately, SNA turned out as a suitable method to govern and quantify the dynamics of ball passing in handball. In addition to traditional performance indicators, it provides an in-depth analysis of passing sequences leading to a better understanding of the nature of the sport and the role of its players.

References

Araújo D, Davids K. Team Synergies in Sport: Theory and Measures. *Frontiers in Psychology*, 2016; 7, 1449

Borgatti SP. Centrality and network flow. *Social networks*, 2005; 27(1): 55-71

Cardinale M, Whiteley R, Hosny AA, Popovic N. Activity Profiles and Positional Differences of Handball Players During the World Championships in Qatar 2015. *International journal of sports physiology and performance*, 2017; 12(7): 908-915

Clemente FM, Martins FML, Kalamaras D, Mendes RS. Network analysis in basketball: Inspecting the prominent players using centrality metrics. *Journal of Physical Education and Sport*, 2015a; 15(2): 212.

Clemente FM, Couceiro MS, Martins FML, Mendes RS. Using network metrics in soccer: a macro-analysis. *Journal of Human Kinetics*, 2015b; 45(1): 123-134

Clemente FM, Martins FML. Network structure of UEFA Champions League teams: association with classical notational variables and variance between different levels of success. *International Journal of Computer Science in Sport*, 2017; 16(1): 39-50

Duch J, Waitzman JS, Amaral LAN. Quantifying the performance of individual players in a team activity. *PloS one*, 2010; 5(6), e10937

Ferguson CJ. An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 2009; 40(5): 532.

Fewell JH, Armbruster D, Ingraham J, Petersen A, Waters JS. Basketball teams as strategic networks. *PloS one*, 2012; 1(11), e47445.

Foretic N, Rogulj N, Papi V. Empirical model for evaluating situational efficiency in top level handball. *International Journal of Performance Analysis in Sport*, 2013; 13(2): 275-293

Freeman LC. Centrality in social networks conceptual clarification. *Social networks*, 1978; 1(3): 215-239

Glazier PS, Davids K. Constraints on the complete optimization of human motion. *Sports Medicine*, 2009; 39(1): 15-28

Gruić I, Vuleta D, Milanović D. Performance indicators of teams at the 2003 world handball championship in Portugal. *Kinesiology*, 2006; 38(2): 164–175

Grund TU. Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, 2012; 34(4): 682-690

Gwet K. *Handbook of inter-rater reliability: How to estimate the level of agreement between two or multiple raters.* Gaithersburg, MD: STATAXIS Publishing Company, 2001

Karcher C, Buchheit M. On-Court demands of elite handball, with special reference to playing positions. *Sports Medicine*, 2014; 44(6): 797-814

Korte F, Lames M. Characterizing different team sports using network analysis. *Current Issues in Sport Science (CISS)*, 2018

Lusher D, Robins G, Kremer P. The application of social network analysis to team sports. *Measurement in physical education and exercise science*, 2010; 14(4): 211-224

Michalsik LB, Aagaard P. Physical demands in elite team handball: comparisons between male and female players. *Journal of Sports Medicine and Physical Fitness*, 2015; 55(9): 878-891

Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*, 2010; 32(3): 245-251

Passos P, Davids K, Araújo D, Paz N, Minguéns J, Mendes J. Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport*, 2011; 14(2): 170-176

Passos P, Araújo D, Volossovitch A. *Performance analysis in team sports*. Taylor & Francis, 53-73; 2016

Pena JL, Touchette H. A network theory analysis of football strategies. *arXiv preprint arXiv:1206.6904*, 2012

Pfeiffer M, Perl J. Analysis of tactical structures in team handball by means of artificial neural networks. *International Journal of Computer Science in Sport*, 2006; 5(1): 4-14

Póvoas SCA, Seabra AFT, Ascensão AAMR, Magalhães J, Soares JMC, Rebelo ANC. Physical and Physiological Demands of Elite Team Handball. *Journal of Strength and Conditioning Research*, 2012; 26(12): 3365–3375

Ramos J, Lopes RJ, Araújo D. What's Next in Complex Networks? Capturing the Concept of Attacking Play in Invasive Team Sports. *Sports Medicine*, 2018; 48(1):17-28

Robinson G, O'Donoghue P. A weighted kappa statistic for reliability testing in performance analysis of sport. *International Journal of Performance Analysis in Sport*, 2007; 7(1): 12-19

Rogulj N, Srhoj V, Srhoj L. The contribution of collective attack tactics in differentiating handball score efficiency. *Collegium Antropologicum*, 2004; 28(2): 739–746

Rogulj N, Vuleta D, Milanovic D, Cavala M, Foretic N. The efficiency of elements of collective attack tactics in handball. *Kinesiologia Slovenica*, 2011; 17(1): 5-14

Srhoj V, Rogulj N, Katić R. Influence of the attack end conduction on match result in handball. *Collegium Antropologicum*, 2001; 25(2): 611–617

Wagner H, Finkenzeller T, Würth S, Von Duvillard SP. Individual and team performance in team-handball: A review. *Journal of Sport Science and Medicine*, 2014; 13(4): 808.

Yamada E, Aida H, Fujimoto H, Nakagawa A. Comparison of game performance among european national women's handball teams. *International Journal of Sport and Health Science*, 2014: 12: 1–10

Figures and Table

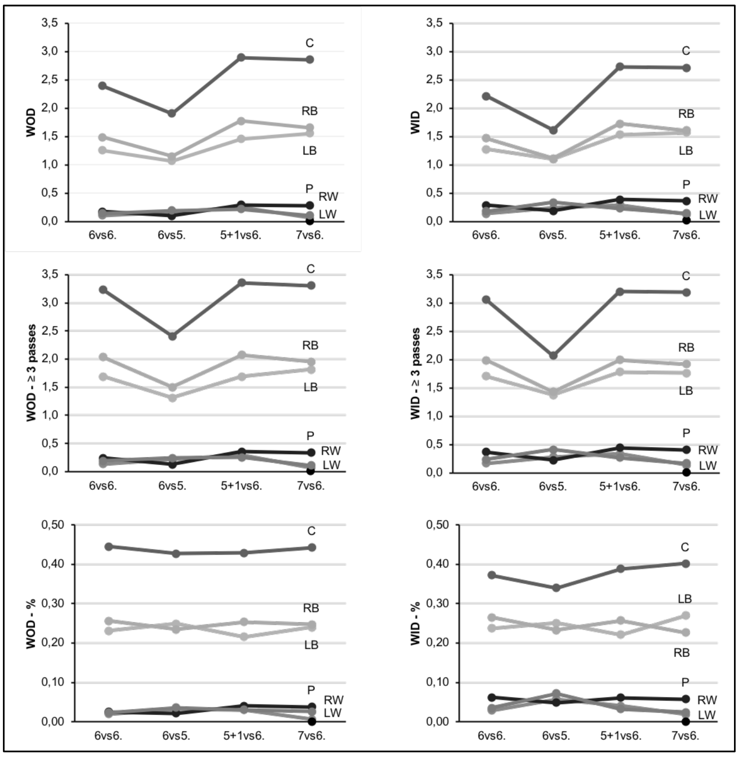

Figure 1: Mean results of C_{WOD} / C_{WID} metrics including ≥ 3 passes and %-values

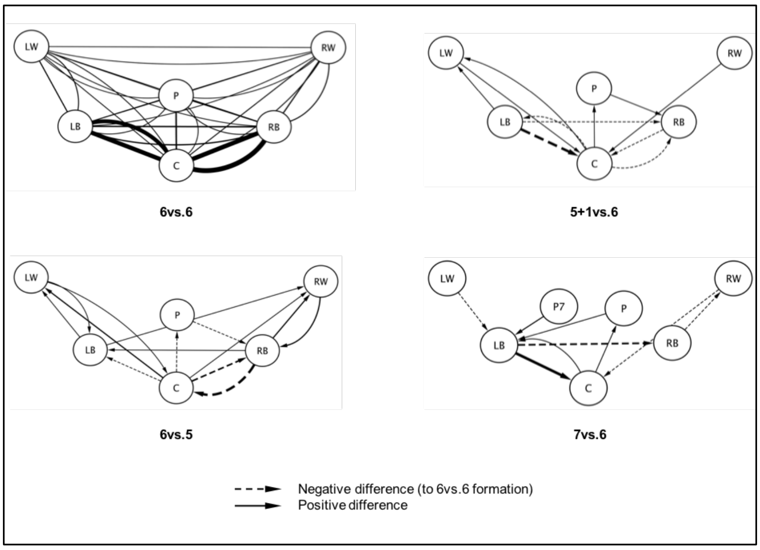

Figure 2: Visualization of passing networks and relative differences between formations

Table 1: Descriptive statistics and post-hoc results for individual metrics

	C	LB	LW	P	P7	RB	RW
C_{WOD}							
6vs.6	2.39 (1.82) all / all	1.25 (1.15) all / 516,76	0.10 (0.34) C,Bs / 65,516	0.17 (0.38) all / all	-	1.49 (1.30) all / 65	0.14 (0.38) C,Bs / 516
6vs.5	1.90 (1.36) all / all	1.07 (0.91) C,P,Ws / 516,76	0.17 (0.42) C,Bs / all	0.10 (0.30) C,Bs / all	-	1.15 (1.01) C,P,Ws / all	0.19 (0.46) C,Bs / -
5+1vs.6	2.89 (1.73) all / 66,65	1.45 (1.09) all / 66,65	0.25 (0.48) C,Bs / 66,65	0.29 (0.48) C,Bs / 66,65	-	1.77 (1.18) all / 65	0.22 (0.46) C,Bs / 66,76
7vs.6	2.85 (1.75) all / 66,65	1.56 (1.05) C,Ps,Ws / 66,65	0.06 (0.23) C,Bs / 65,516	0.28 (0.49) C,Bs / 66,65	0.01 (0.09) C,Bs	1.66 (1.15) C,Ps,Ws / 65	0.11 (0.31) C,Bs / 516
C_{WID}							
6vs.6	2.21 (1.83) all / all	1.27 (1.15) all / 516,76	0.13 (0.38) C,Bs,P / 65,516	0.28 (0.49) C,Bs / 65, 516	-	1.46 (1.28) all / 65,516	0.18 (0.43) C,Bs,P / 65
6vs.5	1.61 (1.27) all / all	1.10 (0.94) C,P,Ws / 516,76	0.23 (0.47) C,Bs / 66	0.19 (0.42) C,Bs / all	-	1.11 (0.99) C,P,Ws / all	0.34 (0.60) C,Bs / all
5+1vs.6	2.73 (1.71) all / 66,65	1.53 (1.12) C,P,Ws / 66,65	0.29 (0.50) C,Bs / 66,76	0.39 (0.52) C,Bs / 66,65	-	1.72 (1.17) C,P,Ws / 66,65	0.23 (0.49) C,Bs / 65
7vs.6	2.71 (1.70) all / 66,65	1.57 (0.97) C,Ps,Ws / 66,65	0.12 (0.33) C,Bs / 516	0.36 (0.50) C,Bs / 65	0.01 (0.09) C,Bs	1.61 (1.24) C,Ps,Ws / 65	0.14 (0.37) C,Bs / 65
C_{FC}							
6vs.6	0.95 (0.03) all / -	0.85 (0.06) C,P,Ws / 76	0.15 (0.08) all / 516	0.29 (0.09) all / -	-	0.86 (0.07) C,P,Ws / -	0.20 (0.09) all / 65
6vs.5	0.94 (0.12) P,Ws / -	0.89 (0.18) P,Ws / -	0.24 (0.23) C,Bs / -	0.24 (0.34) C,Bs / 516	-	0.83 (0.23) P.Ws / 516	0.33 (0.25) C,Bs / 66
5+1vs.6	0.97 (0.08) P,Ws / -	0.88 (0.14) P,Ws / -	0.28 (0.24) C,Bs,P / 66,76	0.41 (0.26) all / 65	-	0.92 (0.10) P,Ws / 65	0.24 (0.21) C,Bs,P / -
7vs.6	0.99 (0.03) Ps,Ws / -	0.96 (0.07) Ps,Ws / 66	0.12 (0.17) C,Bs,Ps / 516	0.43 (0.39) C,Bs,P7,Ws / -	0.02 (0.08) all	0.92 (0.10) Ps,Ws / -	0.21 (0.26) C,Bs,Ps / -
C_{FC3}							
6vs.6	0.94 (0.08) all / -	0.72 (0.12) all / 76	0.12 (0.10) all / -	0.21 (0.11) C,Bs,LW / -	-	0.77 (0.12) all / -	0.18 (0.11) all-P / 65
6vs.5	0.92 (0.20) RB, P, Ws / -	0.82 (0.40) P,Ws / -	0.23 (0.31) C,Bs / -	0.20 (0.36) C,Bs / -	-	0.72 (0.35) C,P,Ws / -	0.33 (0.26) C,Bs / 66, 516
5+1vs.6	0.96 (0.20) all / -	0.76 (0.37) C,P,Ws / 76	0.17 (0.20) C,Bs / -	0.26 (0.17) C,Bs / -	-	0.73 (0.35) C,P,Ws / -	0.13 (0.26) C,Bs / 65
7vs.6	0.97 (0.29) RB,Ps,Ws / -	0.95 (0.32) RB,Ps,Ws / 66,516	0.11 (0.29) C,Bs,P7 / -	0.19 (0.30) C,Bs,P7 / -	0.0 (0.00) all	0.71 (0.38) all / -	0.14 (0.30) C,Bs,P7 / -

Subscripts indicate to which playing positions (part before /) or tactical formation (part after /) given value is statistically different for $p < .05$, e.g. C: given value is statistically different to the value of the center; 66: given value is statistically different to the value in the 6vs.6 formation; All: statistically different to all other playing positions / formations; Bs include LB and RB; Ws include LW and RW; Ps include P and P7

Table 2: Descriptive statistics and post-hoc results for team metrics

	6vs.6	6vs.5	5+1vs.6	7vs.6
C_D	0.17 (0.09) 516	0.17 (0.08) 516	0.22 (0.11) all	0.19 (0.09) 516
C_{WDC}	0.34 (0.07) 516	0.33 (0.08) 76	0.32 (0.08) 66,76	0.35 (0.07) 65,516

Subscripts indicate to which tactical formation given value is statistically different for $p < .05$, e.g. 66: given value is statistically different to the value in the 6vs.6 formation; All: statistically different to all other tactical formations

A.5 Study 3 - Permission for the inclusion in dissertation

Friday, July 26, 2019 at 4:54:33 PM Central European Summer Time

Betreff: Re: Frontiers: Congratulations! Your article is published
Datum: Freitag, 26. Juli 2019 um 16:53:02 Mitteleuropäische Sommerzeit
Von: Frontiers Psychology Production Office
An: Korte, Florian

Good afternoon Dr. Korte,

Thank you for your email.

You can include it in your dissertation. We just ask that you cite the article correctly.

Kind regards,
Folakemi

On Fri, Jul 26, 2019 at 10:47 AM Florian Korte (Via FrontiersIn) <noreply@frontiersin.org> wrote:

> Dear Production Office,
>
> thank you very much for the quick publication process.
>
> Following the Open-Access Policy, I would need a quick confirmation that I can include the published article "Play-by-Play Network Analysis in Football" in my dissertation. If you could please provide me with a short confirmation that this is ok, that would be very helpful. Thank you very much!
>
> Best regards,
>
> Florian Korte

A.6 Study 3 - Original article

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 July 2019
doi: 10.3389/fpsyg.2019.01738

Play-by-Play Network Analysis in Football

Florian Korte, Daniel Link, Johannes Groll and Martin Lames*

Chair of Performance Analysis and Sports Informatics, Technical University of Munich, Munich, Germany

This study identifies dominant and intermediary players in football by applying a play-by-play social network analysis (SNA) on 70 professional matches from the 1. and 2. German Bundesliga during the 2017/2018 season. SNA provides a quantification of the complex interaction patterns between players in team sports. So far, the individual contributions and roles of players in football have only been studied at match-level considering the overall passing of a team. In order to consider the real structure of football, a play-by-play network analysis is needed that reflects actual interplay. Moreover, a distinction between plays of certain characteristics is important to qualify different interaction phases. As it is often impossible to calculate well known network metrics such as betweenness on play-level, new adequate metrics are required. Therefore, flow betweenness is introduced as a new playmaker indicator on play-level and computed alongside flow centrality. The data on passing and the position of players was provided by the Deutsche Fußball Liga (DFL) and gathered through a semi-automatic multiple-camera tracking system. Central defenders are identified as dominant and intermediary players, however, mostly in unsuccessful plays. Offensive midfielders are most involved and defensive midfielders are the main intermediary players in successful plays. Forward are frequently involved in successful plays but show negligible playmaker status. Play-by-play network analysis facilitates a better understanding of the role of players in football interaction.

Keywords: performance analysis, football, temporal networks, flow centrality, intermediary player

OPEN ACCESS

Edited by:
Daniel Vieira Barreira, University of Porto, Portugal

Reviewed by:
Hugo Borges Sarmento, University of Coimbra, Portugal
Gibson Moreira Praça, Federal University of Minas Gerais, Brazil

****Correspondence:***
Florian Korte
korte@cdtm.de

Specialty section:
This article was submitted to Movement Science and Sport Psychology, a section of the journal Frontiers in Psychology

Received: *30 April 2019*
Accepted: *12 July 2019*
Published: *25 July 2019*

Citation:
Korte F, Link D, Groll J and Lames M (2019) Play-by-Play Network Analysis in Football. Front. Psychol. 10:1738. doi: 10.3389/fpsyg.2019.01738

INTRODUCTION

Football teams are described as groups that interact in a dynamic and interdependent way to achieve their common goal (Ribeiro et al., 2017). Understanding the individual role of each player in that dynamic process is highly relevant to uncover how a team operates (Vilar et al., 2013). Although collective behavior within teams is frequently linked to performance outcomes in sports, the impact of individual players on team performance requires further research (Duch et al., 2010). Therefore, identifying methods that offer a quantification of crucial players in the interaction of teams contributes to performance analysis in football.

Social network analysis (SNA) has been identified as a suitable method as it addresses the interdependencies in teams by modeling the interaction based on passes. Passos et al. (2011) describe the potential of SNA by modeling intra-team coordination as the frequent passing interaction taking place between players in team sports. Pena and Touchette (2012) and Grund (2012) build on this idea by connecting network properties to performance outcomes in football. Since then, there has been a growing body of research applying SNA by exploiting passing networks to understand the properties of team performance and the underlying individual contribution of players (Sarmento et al., 2018). The latter is of interest as each player has a specific position to play and role to accomplish in order to contribute to the common goal of winning (Bourbousson et al., 2010). The majority of research in football follows a static analysis on match-level by calculating centrality metrics based on the aggregated passing data in a match. In these studies, the contribution of players to the overall team performance is often described by counting the total number of successfully played and received passes through different degree measures (Clemente et al., 2015; Gama et al., 2015; Trequattrini et al., 2015). Moreover, the intermediary role of players to connect their team mates as bridging players by distributing the ball is frequently assessed by applying betweenness and closeness measures to the overall passing interaction between players across a match (Clemente et al., 2015, 2016a; Aquino et al., 2018; Castellano and Echeazarra, 2019).

Based on these existing studies that apply SNA in football, Ramos et al. (2018) demand a breakdown of the analysis to a play-by-play level to consider the temporal character of football. This implies that passing sequences should be evaluated separately instead of examining the aggregated passing data across a match. Moreover, they emphasize that an analysis on match-level to detect intermediary players through the application of betweenness and closeness measures assumes certain properties about interplay in football that might not be adequate, e.g., the proposition that ball flow follows the shortest paths over the graph which results from the aggregation of all passes in a match. That means that the current approaches do not actually consider the actual sequence of ball passing in order to detect players that are in fact connecting their team members through passing. Instead, the overall intensity of passing across a match is used to approximate bridging players. Third, the authors also suggest a distinction between plays of certain characteristics to ensure a qualitative component to the analysis that bridges the gap between SNA and performance outcomes and fosters the practical impact of the approach.

Some studies already tackle certain aspects of the proposition. Yamamoto and Yokoyama (2011) break down matches in time intervals to meet the temporal character of football. Pina et al. (2017) differentiate between successful and unsuccessful interaction based on aggregated passing networks during certain time intervals. Yet, these approaches do not reflect actual interplay as the analysis is built on aggregated passing data across a number of plays and hence does not consider actual interplay as it unfolds. The reason why most studies conduct an analysis on interval-level instead of play-level is due to the character of plays in football and the current limitations of SNA in sports. In a study by Tenga et al. (2010), 50% of all plays consist of two passes or less and only 20% of all plays take more than four passes. Thus, only a limited number of players are involved in individual plays and it is often not possible to calculate well known individual metrics such as betweenness or closeness on that level of analysis. Moreover, until recently, the regular availability of action feeds in professional football that enable a play-by-play network analysis was limited. In a recent study, Mclean et al. (2018) compute SNA metrics on play-level by analyzing the team interaction properties of goal scoring networks and modeling zones on the playing field as separate nodes to assess how attacks evolve across the pitch. However, there is no differentiation between successful and unsuccessful plays and no assessment of the contribution of individual players.

To summarize, previous research in football has not identified the individual contribution and especially the intermediary role of players based on a separate evaluation of passing sequences. Studies were only executed based on aggregated passing data across time intervals or the entire match. Thus, this study applies and proposes adequate metrics that quantify individual performance on play-level while connecting the results to performance variables. Moreover, a distinction between dominant and intermediary players on play-level is provided. Building on Clemente et al. (2016b), dominant players on match-level are frequently involved in interplay while intermediary players link other teammates during a match.

Following Fewell et al. (2012), flow centrality is calculated to assess the individual dominance on play-level by focusing on the overall involvement during all plays in a match. The intermediary role of players is quantified by counting the share of plays in which the players are actually in-between other teammates. The metric which we call flow betweenness considers the actual sequential pattern of passing and overcomes the issue of short plays, in terms of number of passes, at the same time. We draw a comparison of network metrics between different playing positions as the applied metrics specify and extend the characterization of roles and tasks of players in football. There is also a differentiation between successful and unsuccessful plays by using the entering of the finishing zone as a proxy for goals scored to achieve a rigorous assessment of individual contribution (Tenga et al., 2010). Additionally, the study draws a comparison to the traditional playmaker indicator of weighted betweenness which is computed at match-level. Using a correlation analysis, we can investigate the degree of similarity between flow-based and common match-level metrics and the circumstances in which the results between flow-based metrics differ.

The novelty about this study is twofold. First, it proposes the breakdown of a football match in its sequential order of passing within ball possessions in order to find actual bridging players that are in-between plays. Therefore, our contribution does not lie in the observation of changes in the pattern of interplay across a match but in the consideration of the temporal order of passes within plays to detect actual intermediate players. The second novelty is a comparison between the network metrics of different playing positions in successful and unsuccessful plays to assess their contribution to the team. We focus on the different outcomes of a play, instead of only assessing successful play outcomes such as Mclean et al. (2018) did or relating individual match-level metrics to match outcomes which accepts potential noise in the analysis. Flow-based metrics quantify the proportional prevalence or intermediary role of players in a match. They appear most fitting in a football context as they are robust to the short plays in football, allow a consideration of the temporal order of passing as proposed through flow betweenness and offer a suitable connection to performance outcomes on play-level.

MATERIALS AND METHODS

Samples

A total of 70 matches between 35 professional male football teams from the 1. and 2. German Bundesliga were analyzed during the 2017/2018 season. Matches were randomly selected from a pool and, on average, teams were present in four matches with no repetition of any encounter. The final sample consists of 24,990 passes captured in 5409 plays.

Procedure

The focus of this study lies on an analysis at play-level. This means that interplay in each ball possession is examined separately instead of evaluating an aggregated passing matrix at match-level. The data was provided by the Deutsche Fußball Liga (DFL). It contains positional data for each player and the ball, which was collected by the multiple-camera tracking system TRACAB® operating at 25 Hz. The validity and reliability of the system was secured in an independent study (Linke et al., 2019). Action feeds including information on passing were also provided and their reliability secured by the DFL. Definitions and validation procedures can be found in the DFL definitions catalog for official match data (2014). Twenty-eight percent of the original data is dropped in the cleaning process providing 8897 plays that clearly identified each ball possession and player involved. In order to conduct our analysis, we capture each play in a two-dimensional passing array consisting of the players in possession of the ball and an index reflecting the sequential order of the ball passing during the play. We also build a corresponding adjacency matrix for each play which are then aggregated across a match to calculate the traditional playmaker indicator on match-level. **Figure 1** provides an example of a passing sequence with its corresponding passing array and adjacency matrix. For the purpose of our study, the final sample (61% of all plays) focuses on plays of at least two completed passes (the minimum play size for having an intermediate player).

We categorize a possession as successful when a team enters the finishing zone, which is a common proxy for goals scored (Tenga et al., 2010). This category includes all plays of at least two passes that lead to entering the finishing zone and sequences are captured until the moment of success (Pina et al., 2017). A play is declared as unsuccessful if ball possession is lost by any means before entering the zone. Neutral plays already start in the finishing zone or consist of set-plays directly entering it. The possession outcome was classified combining the positional data provided for each player and the ball with the information on the standardized pitch sizes in the German Bundesliga and dimensions of the finishing zone as defined by Tenga et al. (2010). The information jointly enabled an automatic evaluation on whether the player in ball possession entered the finishing zone or whether a successful pass was played to a teammate in that designated area. That way, we could also detect whether a possession starts in the finishing zone in order to declare it as neutral. This leads to 21.5% successful plays, 74.5% are declared unsuccessful and a remainder consisting of 4% in neutral plays.

Playing positions are tracked to facilitate an evaluation of the individual contribution of players in our study. Multiple players may be assigned to the same tactical position. Average metric values are reported to evaluate the performance of the playing positions in this case. The final classification is in line with previous studies focusing on players in football (Clemente and Martins, 2017; Korte and Lames, 2018). We codify the following seven playing positions according to the definitions catalog for official match data provided by the DFL (2014): (i) goalkeeper (GK); (ii) central defender (CD); (iii) external defender (ED); (iv) central defensive midfielder (CDM); (v) external midfielder (EM); (vi) central offensive midfielder (COM); and (vii) forward (F). Substitutions are handled through a reassignment of playing positions according to the DFL data provided. By codifying playing positions, in comparison to specific player tracking, there is no need to standardize the obtained values according to time on the field (Praça et al., 2019).

Network Metrics

The analysis was carried out using the Python package NetworkX® and the software libraries pandas and NumPy. A set of individual metrics was computed to achieve a quantification of the contribution of playing positions in a team's interplay. By calculating flow centrality, a concept first introduced in basketball by Fewell et al. (2012), we capture the involvement of each playing position in all plays across a match. Building on this metric and random-walk betweenness by Newman (2005), we also compute a new metric called flow betweenness. For comparison purposes, we also calculate weighted betweenness scores for each playing position based on the aggregated passing data across a match. Whereas the two play-level metrics model pass interactions as walks, the weighted betweenness computation is based on the concept of shortest paths to evaluate the intermediary role of players (Ramos et al., 2018).

Flow Centrality

For each player, flow centrality measures the fraction of plays (or attack units) that it is involved in at least once relative to all plays by its team. Thus, an indication on the overall involvement of all playing positions across a match is provided. Following Fewell et al. (2012), flow centrality index, $C_{\mathrm{FC}}(n_i)$, for player i is calculated as,

$$C_{\mathrm{FC}}(n_i) = \frac{\sum_{k=1}^{m} p_k(n_i)}{M} \quad (1)$$

where M denotes the total number of plays by a team in a match and $p_k(n_i)$ denotes the k-th play in which n_i is part of at least once. By construction, flow centrality values are bounded between 0 and 1. The extreme value of 0 signals that a player was not part of any play in terms of passing or receiving the ball. A value of 1 means that a player was at least involved once in every play of its team during the match. Any flow centrality value in between can be interpreted as the proportion of plays that a player was involved in relative to all plays by its team.

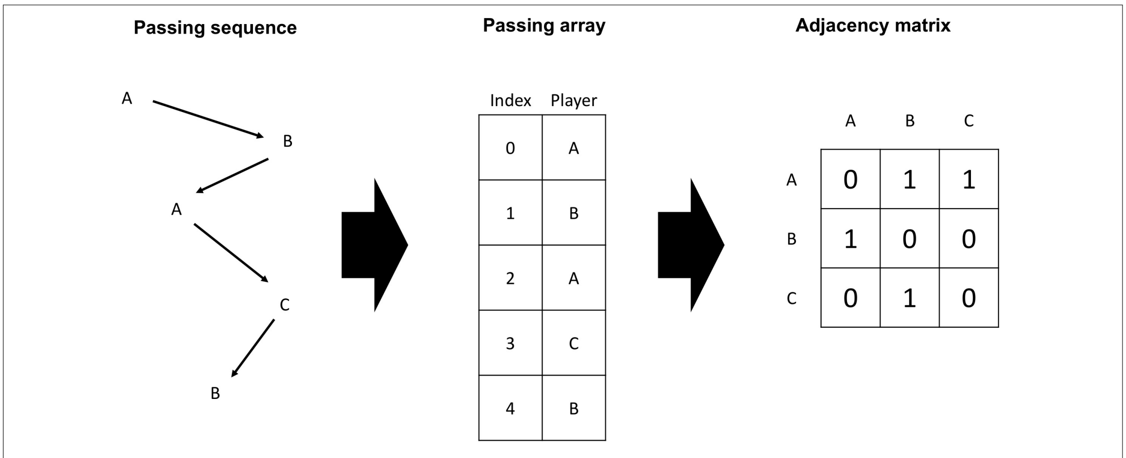

FIGURE 1 | Example of a passing sequence with its corresponding passing array and adjacency matrix.

Flow Betweenness

For each player, flow betweenness measures the fraction of plays in which it functions as an intermediary player relative to all plays by its team. We define a player as intermediate in a play if it actually functions as a bridging player in terms of passing between any other two players. Flow betweenness index, $C_{FB}(n_i)$, for player i is calculated as,

$$C_{FB}(n_i) = \frac{\sum_{k=1}^{m} b_k(n_i)}{M} \tag{2}$$

where M denotes the total number of plays by a team in a match and $b_k(n_i)$ denotes the k-th play in which n_i is functioning as an intermediary player. In contrast to C_{FC}, which only tracks involvement, C_{FB} considers the actual passing sequence of a play to track whether a player is positioned in between a sequence to function as a bridging unit. Flow betweenness values are also bounded between 0 and 1. Values of 0 signal that a player did not once receive the ball by a teammate and successfully passed it on to another teammate in any play during a match. A value of 1 means that a player received and passed on the ball at least once in every play of its team. Values in between the extreme values are again the proportion of plays that a player functioned in as a bridging unit relative to all plays by its team.

While being *in-between* always implies being *involved* in a play, the reversal is not true. Initiating or being at the end of a play implies that a player is *involved* but not *in-between* a ball possession. Therefore, the flow centrality value of a player in a match is always at least as high as its corresponding flow betweenness value.

Weighted Betweenness

Weighted betweenness assesses how often a player is in-between any other two players of its team measured by their strongest passing connections across a match. Thus, its betweenness character is built on aggregated match data and does not necessarily imply that the player functioned as a bridging unit within plays. It is often used as a playmaker indicator (Pena and Touchette, 2012; Clemente and Martins, 2017). The weighted betweenness index, $C_{WB}(n_i)$, for player i is calculated as,

$$C_{WB}(n_i) = \sum_{j \neq k \neq i} \frac{g_{jk}^{i}}{g_{jk}} \tag{3}$$

where g_{jk}^{i} is the number of strongest passing connections via player i from players j to k and g_{jk} the total number of strongest passing connections between players j and k. The values of weighted betweenness are bounded between 0 and 1 reflecting the proportion of strongest passing connections between any two players in the network that lead via a particular player.

Statistical Procedures

Data were analyzed for normality using Shapiro–Wilk tests. Since only 40% of data was normally distributed, non-parametric statistical analyses were used.

For both play-level metrics, multiple Kruskal–Wallis H test are executed to test for statistical differences between playing positions for the entire sample.

In order to differentiate between successful and unsuccessful plays, we apply Kruskal–Wallis H tests on two separate samples, filtering for successful and unsuccessful plays accordingly, to detect differences in play-level metrics between playing positions. Moreover, multiple Mann–Whitney U tests are conducted for each playing position to investigate statistical differences in metrics between the different play outcomes.

As the share of successful plays is severely higher in plays starting from the opponent's half than from the own half of a team (28.3–16.8%), we suspect the starting half to be a moderator variable that could partly influence differences in involvement in successful against unsuccessful plays across playing positions. Hence, the same procedure to differentiate between successful and unsuccessful plays is repeated focusing on plays starting from a team's own half. For each approach, Dunn-Bonferroni post-hoc tests offer pairwise comparisons between groups, respectively.

Our statistical analysis is conducted at a 5% significance level. Following Ferguson (2009) and Cohen (2008), non-parametric estimates of η^2 are reported to interpret the effect size according to the following criteria: no effect ($\eta^2 < 0.04$); small effect ($0.04 \leq \eta^2 < 0.25$); moderate effect ($0.25 \leq \eta^2 < 0.64$); strong effect ($\eta^2 \geq 0.64$). Ninety percentage confidence intervals for η^2 are calculated following Hopkins (2017).

To assess the relationship between the network metrics, a correlation analysis is carried out across the sample. First, the Pearson correlation coefficients between C_{FC} (C_{FB}) and C_{WB} are calculated, respectively to evaluate the association between metrics conducted on play-level and match-level. Second, the Pearson correlation coefficient between C_{FC} and C_{FB} is computed to assess differences between the two metrics. By construction of C_{FB}, we expect the metric to be dependent on the number of passes per play. Therefore, coefficients for three subsets are calculated, following Tenga et al. (2010): (i) matches with on average less than three passes per play; (ii) matches with three to five passes per play; and (iii) matches with more than five passes per play. The strength of the correlation is assessed according to the following guide by Evans (1996): moderate ($0.40 \leq r < 0.60$); strong ($0.60 \leq r < 0.80$); very strong ($0.80 \leq r < 1.0$). Ninety-five percentage confidence intervals for r are calculated following Hopkins (2017).

RESULTS

General Analysis

We find significant differences between playing positions for $\left(p < 0.001; \eta^2 = 0.23, CI[0.12, 0.34], \text{ small effect}\right)$ and $C_{FB}(p < 0.001; \eta^2 = 0.34, CI\,[0.17, 0.51], \text{ moderate effect})$. **Figure 2** shows that CDs are significantly more involved (47% of all plays) and also function more often as intermediators (28%) in a match than any other tactical position. Fs are least involved in plays (28%) and take on an intermediary role in 13% of all attack units. By definition of the metrics, the C_{FB} value is lower for each playing position than its corresponding C_{FC} value. The largest difference between both metrics is reported for the GK.

Success Analysis

Table 1 presents the results of the Kruskal–Wallis H tests for C_{FC} and C_{FB} differentiating between successful and unsuccessful plays of the overall sample and focusing on plays starting from a team's own half. All eight tests reveal statistically significant differences between playing positions for the respective subsample with varying effect sizes.

Table 2 presents the results of the Mann–Whitney U tests for each flow-based metric, playing position and differentiating also between the overall sample and focusing on plays starting from a team's own half. Apart from the ED position, the tests reveal significant differences between successful and unsuccessful plays in terms of C_{FC} and C_{FB} for all other playing positions. However, some effect sizes are small to negligible.

In general, offensive positions (EMs, COMs, Fs) are significantly more involved in successful than in unsuccessful plays, whereas defensive positions (GK, CDs) are significantly less involved in successful plays. The C_{FC} and C_{FB} values per playing position for each play outcome and the results of the post-hoc tests can be taken from **Table 3**. COMs have the highest involvement in successful plays (50%) while GK take only part in 17% of all successful plays. CDs are not only most prevalent in unsuccessful plays (51%), followed by GK and EDs, but are also in-between most unsuccessful plays (37%). In contrast, CDMs are the leading intermediary players (32% of all successful plays), while GK and Fs have the lowest values in this category.

Figure 2 shows that the difference between both metric scores is increasing as more offensive the playing position is on the pitch for successful plays. Moreover, while defenders and defensive midfielders are functioning as bridging players in 70–75% of all plays they are involved in, the shares for GK and Fs are only 40–50%.

Focusing on plays starting in the own half of a team, the difference of involvement and the intermediary role between successful and unsuccessful plays is reported smaller for defensive positions in comparison to the results of all plays. This indicates that the significantly large gap is moderated by the starting half of a play. In comparison to the analysis on all plays, EMs are most involved in successful plays starting from its team's own half and come level with the intermediary player values of the other midfield positions.

Correlation Analysis

The Pearson correlation coefficients between each flow-based metric and the weighted betweenness scores on match-level indicate a strong positive relationship for C_{FC} $(r = 0.68;\ CI[0.64,\ 0.72];\ p < 0.001)$ and for C_{FB} $(r = 0.67;\ CI[0.63,\ 0.71];\ p < 0.001)$. The correlation coefficient between the involvement and intermediary metric on play-level indicates a very strong positive relationship at first sight $(r = 0.89;\ CI[0.87,\ 0.90];\ p < 0.001)$. However, the correlation strength depends on the average number of passes in plays during a match. Whereas we find a very strong positive relationship in matches with more than five passes on average per play $(r = 0.95;\ CI[0.93,\ 0.96];\ p < 0.001)$ and also in matches with three to five passes per play $(r = 0.86;\ CI[0.84,\ 0.88];\ p < 0.001)$, there is only a moderate positive relationship in matches with less than three passes per play $(r = 0.56;\ CI[0.38,\ 0.70];\ p < 0.001)$.

DISCUSSION

The study reveals statistical significance between playing positions in successful and unsuccessful plays in football with regard to flow centrality and the newly introduced flow betweenness. Moreover, for the majority of playing positions there are significant differences between play outcomes with regard to both flow-based metrics. Effect sizes found were small to moderate with regard to playing positions and mostly small in terms of play outcomes.

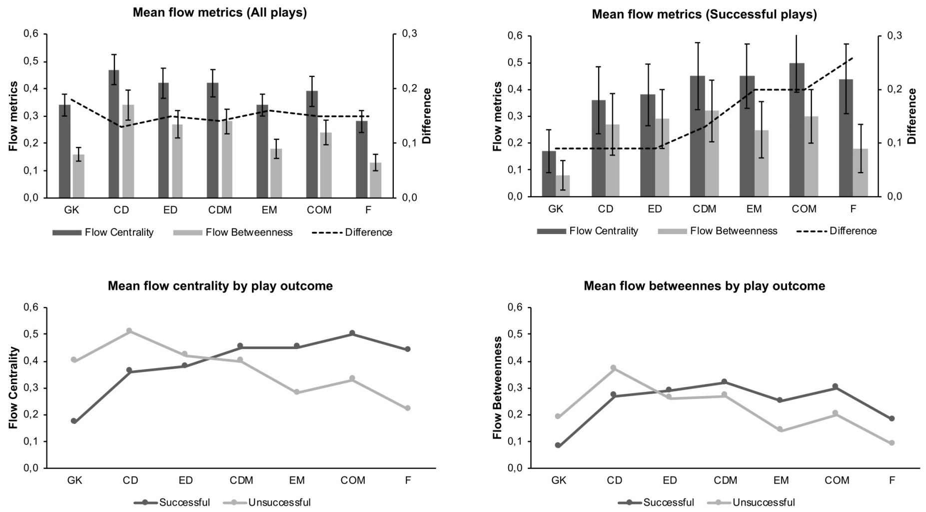

FIGURE 2 | Mean results of flow-based metrics by playing position.

TABLE 1 | Kruskal–Wallis *H* test results for playing position comparison per play outcome.

	Successful plays				Unsuccessful plays			
	H	p	η^2	CI of η^2	H	p	η^2	CI of η^2
C_{FC}								
All plays	172.77	<0.001	0.12	[0.06, 0.18]	509.25	<0.001	0.37	[0.19, 0.55]
Own half	35.3	<0.001	0.03	[0.02, 0.04]	455.77	<0.001	0.39	[0.20, 0.39]
C_{FB}								
All plays	164.83	<0.001	0.12	[0.06, 0.18]	571.58	<0.001	0.42	[0.21, 0.63]
Own half	63.34	<0.001	0.05	[0.03, 0.07]	422.16	<0.001	0.36	[0.18, 0.54]

Overall involvement and the frequency of being an intermediary player is lower in successful than unsuccessful plays for defensive playing positions and the other way around for offensive positions. This turns out to be partly moderated by the origin of play on the pitch, which is incident to differences in success probability. Besides, the results offer first insights into the differences between dominant and intermediary players in football measured by the two play-level metrics.

While our analysis presents CDs as the most involved and intermediary playing position, most studies traditionally ascribe midfielders the most dominant and intermediary role in football (Cotta et al., 2013; Clemente et al., 2015, 2016b). There is also literature that positions forward (Clemente et al., 2016a) and EDs (Gama et al., 2014) as intermediary players. There are multiple reasons why our results differ from past studies aside from the fact that a different sample was considered.

First, involvement (or dominance) in interplay in football is often measured by the number of successfully played and received passes in a match in form of weighted in-degree and weighted out-degree (Clemente et al., 2016a). However, there is no information on whether the passes occurred in a limited amount of longer plays, in terms of number of passes, or frequently across a match. This implies that players with high flow centrality do not necessarily play and receive most passes during a match but are most frequently part of plays across an entire match. Therefore, the match-level metrics measure the share in a team's total passing while the play-level metric evaluates the prevalence in plays across a match.

Second, intermediary players in football, often referred to as playmakers, have formerly been determined by how often they are on average the strongest connector between the other players based on the aggregated passing data of a match (Trequattrini et al., 2015; Arriaza-Ardiles et al., 2018). However, that does not imply that the player frequently distributed the ball between other players. In an extreme scenario, a midfielder who frequently loses a ball received by defenders and frequently wins balls from the opponent and passes it to forward positions is identified as a bridging player without ever actually connecting defense and offense during a play. Flow betweenness detects how often a player is actually in-between two other players during a play and is in fact acting as an intermediary player.

TABLE 2 | Mann–Whitney *U* test results for play outcome comparison per playing position.

	Play outcome						
	GK	**CD**	**ED**	**CDM**	**EM**	**COM**	**F**
C_{FC}							
All plays							
H	2516.5	30081	11691	31305	4070	2879	11747
p	<0.001	<0.001	0.090	<0.001	<0.001	<0.001	<0.001
η^2	0.41	0.12	0.01	0.02	0.18	0.2	0.24
CI of η^2	[0.21, 0.61]	[0.06, 0.18]	[0.00, 0.02]	[0.01, 0.03]	[0.09, 0.27]	[0.10, 0.30]	[0.12, 0.36]
Own half							
H	3522	29847.5	8710	21863.5	2844.5	3366.5	8455.5
p	<0.001	<0.001	0.200	0.002	<0.001	0.002	<0.001
η^2	0.17	0.02	0.01	0.02	0.21	0.04	0.23
CI of η^2	[0.09, 0.25]	[0.01, 0.03]	[0.00, 0.02]	[0.01, 0.03]	[0.11, 0.31]	[0.02, 0.06]	[0.12, 0.34]
C_{FB}							
All plays							
H	4129	35378.5	12186	32642.5	5275	3888.5	20974.5
p	<0.001	<0.001	0.230	0.005	<0.001	<0.001	<0.001
η^2	0.24	0.07	0.01	0.01	0.08	0.09	0.04
CI of η^2	[0.12, 0.36]	[0.04, 0.10]	[0.00, 0.02]	[0.00, 0.02]	[0.04, 0.12]	[0.05, 0.13]	[0.02, 0.06]
Own half							
H	3723.5	30416	8928	21868	4082.5	3685.5	16594.5
p	<0.001	<0.001	0.310	0.002	<0.001	0.025	0.017
η^2	0.15	0.02	0.01	0.02	0.08	0.02	0.01
CI of η^2	[0.08, 0.22]	[0.01, 0.03]	[0.00, 0.03]	[0.01, 0.03]	[0.04, 0.12]	[0.01, 0.03]	[0.00, 0.02]

Following Fewell et al. (2012) and Ramos et al. (2018), an evaluation based on the involvement in plays, however, becomes considerably more useful when making a distinction between plays with certain characteristics. While CDs appear to be the dominant and intermediary players, the majority of plays they are part of do not enter the finishing zone. In contrast, COMs are most often part of successful plays and CDM is the most intermediary position in these situations. In general, defensive playing positions show a higher involvement in unsuccessful than successful plays. Focusing solely on plays that originate in the own half of a team offsets that difference to a certain extent. Similar to previous studies (Tenga et al., 2010; Mclean et al., 2018), the share of successful plays was higher for plays starting in the opposite half and, thus, involved more offensive playing positions. The analysis on plays starting in the own half of a team partly neutralized this imbalance. This is reflected in the small to negligible effect sizes obtained when evaluating the differences in flow-based metrics between playing positions focusing on successful plays starting from the own half. Moreover, the effect sizes for differences between successful and unsuccessful plays decreases for defensive playing positions. Apart from that, the analysis provides an insight into how attacks from a team's own half are most frequently structured. The increased metric values of the EM position in contrast to the analysis on the total sample suggest that plays were frequently build via wing positions. Therefore, the approach of subdividing the sample into different types of plays with different outcomes provides a certain quality to the analysis that goes beyond pure prevalence in plays by offering a richer insight into the structure of plays in different contexts.

The distinction between being involved and acting as an intermediary player is recognizable when focusing the analysis on successful plays. From a pure descriptive perspective, the more offensive the playing position is located on the pitch the higher its difference between the two play-level metrics. Offensive players such as forward are often involved in successful plays, however, not in order to distribute the ball but rather to take on the role of finishing attacks. While the absolute difference between the flow-based metrics for GKs might be small, the share of plays in which they function in-between others measured against all plays they are involved in is quite low. Their task is often that of an initiator of plays rather than being a bridging player. Therefore, their intermediary status is relatively low. In contrast, CDMs are similarly often involved in successful plays as forward but have a substantially higher share of incidences in which they function as a bridging player at the same time.

The correlation analysis underlines the insights of our study, especially that (i) different results on playmakers in football might be obtained when substituting match-level with play-level metrics and (ii) a distinction between play-level metrics is necessary as they emphasize different tasks among playing positions. Ramos et al. (2018) first suggested that flow centrality might be a suitable playmaker indicator that highlights intermediary players on play-level to replace the average-based analysis provided by weighted

TABLE 3 | Descriptive statistics and post-hoc results of C_{FC} and C_{FB}.

	GK	CD	ED	CDM	EM	COM	F
C_{FC}							
All plays							
General	0.34 (0.08)^{b,c,d,g}	0.47 (0.11)^{all}	0.42 (0.11)^{a,b,e,g}	0.42 (0.10)^{a,b,e,g}	0.34 (0.08)^{b,c,d,g}	0.39 (0.11)^{b,g}	0.28 (0.08)^{all}
Successful	0.17 (0.16)$^{all/yes}$	0.36 (0.25)$^{a,d,e,f/yes}$	0.38 (0.23)$^{a,f/no}$	0.45 (0.25)$^{a,b/yes}$	0.45 (0.24)$^{a,b/yes}$	0.50 (0.22)$^{a,b,c/yes}$	0.44 (0.26)$^{a/yes}$
Unsuccessful	0.40 (0.12)$^{b,e,f,g/yes}$	0.51 (0.15)$^{all/yes}$	0.42 (0.14)$^{b,e,f,g/no}$	0.40 (0.13)$^{b,e,f,g/yes}$	0.28 (0.12)$^{a,b,c,d,g/yes}$	0.33 (0.12)$^{a,b,c,d,g/yes}$	0.22 (0.11)$^{all/yes}$
Own half							
Successful	0.29 (0.27)$^{b,d,e,f,g/yes}$	0.43 (0.32)$^{a/yes}$	0.38 (0.31)$^{e/no}$	0.44 (0.31)$^{a/yes}$	0.52 (0.32)$^{a,c/yes}$	0.42 (0.31)$^{a/yes}$	0.47 (0.32)$^{a/yes}$
Unsuccessful	0.48 (0.15)$^{d,e,f,g/yes}$	0.52 (0.17)$^{c,d,e,f,g/yes}$	0.41 (0.15)$^{b,e,f,g/no}$	0.37 (0.15)$^{a,b,e,g/yes}$	0.27 (0.13)$^{a,b,c,d,g/yes}$	0.32 (0.13)$^{a,b,c,g,f/yes}$	0.20 (0.12)$^{all/yes}$
C_{FB}							
All plays							
General	0.16 (0.05)^{b,c,d,f}	0.34 (0.11)^{all}	0.27 (0.10)^{a,b,e,g}	0.28 (0.09)^{a,b,e,f,g}	0.18 (0.07)^{b,c,d,f,g}	0.24 (0.09)^{a,b,d,e,g}	0.13 (0.06)^{b,c,d,e,f}
Successful	0.08 (0.11)$^{all/yes}$	0.27 (0.23)$^{a,g/yes}$	0.29 (0.22)$^{a,g/no}$	0.32 (0.23)$^{a,g/yes}$	0.25 (0.21)$^{a,g/yes}$	0.30 (0.20)$^{a,g/yes}$	0.18 (0.18)$^{all/yes}$
Unsuccessful	0.19 (0.09)$^{b,c,d,e,g/yes}$	0.37 (0.15)$^{all/yes}$	0.26 (0.13)$^{a,b,e,f,g/no}$	0.27 (0.13)$^{a,b,e,f,g/yes}$	0.14 (0.09)$^{a,b,c,d,f/yes}$	0.20 (0.11)$^{b,c,d,e,g/yes}$	0.09 (0.09)$^{a,b,c,d,f/yes}$
Own half							
Successful	0.13 (0.19)$^{b,c,d,e,f/yes}$	0.32 (0.29)$^{a,g/yes}$	0.31 (0.30)$^{a/no}$	0.34 (0.29)$^{a,g/yes}$	0.33 (0.30)$^{a,g/yes}$	0.29 (0.26)$^{a/yes}$	0.22 (0.27)$^{b,d,e/yes}$
Unsuccessful	0.23 (0.11)$^{b,e,g/yes}$	0.37 (0.16)$^{all/yes}$	0.26 (0.14)$^{b,c,e,f,g/no}$	0.26 (0.14)$^{b,d,e,f,g/yes}$	0.13 (0.10)$^{a,b,c,d,f/yes}$	0.19 (0.11)$^{b,c,d,e,g/yes}$	0.10 (0.09)$^{a,b,c,d,f/yes}$

Significantly different to GK^a, CD^b, ED^c, CDM^d, EM^e, COM^f, F^g, and all at $p<0.05$; yes/no indicates whether value is statistically different to other play outcome at $p<0.05$.

betweenness on match-level. However, the relationship between both metrics does not suggest that the same matter is measured.

Differences between values of flow centrality and flow betweenness for playing positions are also confirmed in the correlation analysis. The overestimated intermediary role of players when simply looking at involvement instead of their in-between positioning in plays is connected with the average number of passes in plays. Shorter plays offer less situations for players to be in-between plays and, thus, a sole involvement measure might exaggerate the intermediary task of a player. Hence, flow betweenness might be a more adequate playmaker indicator.

In general, the play-by-play network analysis approach allows a more contextualized performance analysis as the role of players in passing sequences of different characteristics can be evaluated separately. Barreira et al. (2015) find that team dynamics are influenced by situational variables such as match status and halves of the match. Controlling for such variables can offer a better understanding of the involvement and intermediary role of players in specific play situations.

Our study also faces some limitations that should be addressed. First, the sample only originates from two professional football leagues and, therefore, the generalizability of our results might be limited. The concern is partly offset by the findings of Mclean et al. (2017) who do not detect significant differences in passing networks between the 2016 European football championships and COPA America football championships.

Second, the determination of playing positions might contrast the less static interpretation of roles in modern football. As we break down the analysis to individual plays, the fixed assignment of positions across a match is even more challenging. We acknowledge the occupation of different areas on the pitch and fulfilling a variety of tasks as part of the role repertoire of playing positions (Korte and Lames, 2018). Hence, the spread in metric values of some playing positions might be ascribed to the mixed role interpretation of players. However, we should stress that playing positions might be interpreted differently not only across matches but also during different phases of a match depending on the specific constraints that players face. This was not considered in the present study.

Third, this study only focuses on plays with at least two completed passes to offer a calculation of flow betweenness across all plays. A study including plays with only one pass would increase the difference between flow centrality and flow betweenness simply because it offers no in-between situations for players. In fact, the correlation coefficient between both metrics greatly decreases ($r = 0.69$) when adding plays with only one pass to the analysis. However, the weakened relationship based on plays of any length also validates the introduction of a new playmaker indicator to reflect the real structure of football on a play-by-play level.

Moreover, it should be stressed that the comparison between successful and unsuccessful plays per playing position could be partly confounded by the cutoff of the passing sequences once the finishing zone is entered. Successful plays continued on average for 0.5 passes after the outcome determination. However, a separate analysis based on the entire passing sequences shows that a wider gap in play involvement between successful and unsuccessful plays for COMs and Fs is the only substantial change.

In addition, the opponent's strength and especially defensive actions were not considered in this study, which could potentially have an impact on the involvement of certain playing positions. Focusing on the attacking side, it should be mentioned that we did not concentrate on identifying different game styles but rather aimed at emphasizing the different roles and contributions of playing positions.

CONCLUSION

This is the first study that performs a play-by-play network analysis in football differentiating between plays of certain characteristics. Moreover, a novel metric is introduced to assess playmakers on play-level as an alternative or extension to flow centrality. Only a limited connection with traditional playmaker indicators on match-level can be detected. Hence, it offers new insights and a better understanding of the roles of playing positions during plays in football.

Central defenders are identified as dominant and intermediary players, however, mostly in unsuccessful plays. COMs are most involved and CDMs function mostly as intermediary players in successful attacks. Fs are frequently involved in successful plays but take on a minor intermediary role.

The practical impact of this study is twofold. First, a playmaker indicator that focuses on actual passing sequences rather than averages across a game was applied to adequately reflect interplay in football. Second, the study provides a more sophisticated understanding of the involvement and role of players in different play situations. Apart from considering play outcome, the play-by-play network analysis approach allows the inclusion of additional situational variables that are relevant to performance in football. The insights and approach of this study could be used and applied in practical performance analysis. By tracking specific players rather than playing positions, clubs can gain a better understanding of the involvement and intermediary role of their individual players in the interplay of the team.

Future studies should continuously focus on developing new SNA-metrics that reflect actual interplay and study the impact of the opponent team on the interaction of the team in ball possession. Moreover, position-specific performance indicators could complement the current play-level approach that solely focuses on whether the finishing zone was reached.

DATA AVAILABILITY

The raw datasets for this manuscript are not publicly available because data sets were collected by Deutsche Fußball Liga (DFL). Requests should be directed to FK, korte@cdtm.de.

ETHICS STATEMENT

Since each player agreed to the video recording of matches on signing their player license, an ethics approval was not required as per applicable institutional and national guidelines. Nevertheless, all procedures performed in the study were in strict accordance with the Declaration of Helsinki as well as with the ethical standards of the local ethics committee.

AUTHOR CONTRIBUTIONS

FK conceived the design, was responsible for the statistical procedures and interpretation of the data, wrote the manuscript, and reached the final version of the manuscript. ML supported the development of the research design and the statistical procedures, assisted in the interpretation of the data, and reviewed the manuscript. JG preprocessed the data and supported the analysis of the matches. DL supported the preprocessing of the data and provided reviews to the manuscript.

FUNDING

This work was supported by the German Research Foundation (DFG) and the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program.

REFERENCES

Aquino, R., Carling, C., Vieira, L., Martins, G., Jabor, G., Machado, J., et al. (2018). Influence of situational variables, team formation and playing position on match running performance and social network analysis of Brazilian professional soccer players. *J. Strength Cond. Res.* doi: 10.1519/JSC.0000000000002725 [Epub ahead of print].

Arriaza-Ardiles, E., Martín-González, J. M., Zuniga, M. D., Sánchez-Flores, J., De Saa, Y., and García-Manso, J. M. (2018). Applying graphs and complex networks to football metric interpretation. *Hum. Mov. Sci.* 57, 236–243. doi: 10.1016/j.humov.2017.08.022

Barreira, D., Garganta, J., Castellano, J., Machado, J., and Anguera, M. T. (2015). How elite-level soccer dynamics has evolved over the last three decades? Input from generalizability theory. *Cuadernos de Psicol. del Deporte* 15, 51–62. doi: 10.4321/s1578-84232015000100005

Bourbousson, J., Poizat, G., Saury, J., and Seve, C. (2010). Team coordination in basketball: description of the cognitive connections among teammates. *J. Appl. Sport Psychol.* 22, 150–166. doi: 10.1080/10413201003664657

Castellano, J., and Echeazarra, I. (2019). Network-based centrality measures and physical demands in football regarding player position: is there a connection? A preliminary study. *J. Sports Sci.* doi: 10.1080/02640414.2019.1589919 [Epub ahead of print].

Clemente, F. M., and Martins, F. M. L. (2017). Who are the prominent players in the UEFA champions league? An approach based on network analysis. *Walailak J. Sci. Technol.* 14, 627–636.

Clemente, F. M., Martins, F. M. L., Wong, P. D., Kalamaras, D., and Mendes, R. S. (2015). Midfielder as the prominent participant in the building attack: a network analysis of national teams in FIFA World Cup 2014. *Int. J. Perform. Anal. Sport* 15, 704–722. doi: 10.1080/24748668.2015.11868825

Clemente, F. M., Martins, F. M. L., and Mendes, R. S. (2016a). Analysis of scored and conceded goals by a football team throughout a season: a network analysis. *Kinesiology* 48, 103–114. doi: 10.26582/k.48.1.5

Clemente, F. M., Silva, F., Martins, F. M. L., Kalamaras, D., and Mendes, R. S. (2016b). Performance analysis tool for network analysis on team sports: a case study of FIFA soccer world cup 2014. *Proc. Ins. Mech. Eng. Part P J. Sports Eng. Technol.* 230, 158–170. doi: 10.1177/1754337115597335

Cohen, B. H. (2008). *Explaining Psychological Statistics*. Hoboken, NJ: John Wiley & Sons.

Cotta, C., Mora, A. M., Merelo, J. J., and Merelo-Molina, C. (2013). A network analysis of the 2010 FIFA world cup champion team play. *J. Syst. Sci. Compl.* 26, 21–42. doi: 10.1007/s11424-013-2291-2

DFL (2014). *Definitionskatalog Offizielle Spieldaten (Definitions for Official Game Data)*. DFL: Frankfurt.

Duch, J., Waitzman, J. S., and Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PLoS One* 5:e10937. doi: 10.1371/journal.pone.0010937

Evans, J. D. (1996). *Straightforward Statistics for the Behavioral Sciences*. Belmont, CA: Thomson Brooks/Cole Publishing Co.

Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. *Prof. Psychol.* 40, 532–538. doi: 10.1037/a0015808

Fewell, J. H., Armbruster, D., Ingraham, J., Petersen, A., and Waters, J. S. (2012). Basketball teams as strategic networks. *PLoS One* 7:e47445. doi: 10.1371/journal.pone.0047445

Gama, J., Couceiro, M., Dias, G., and Vaz, V. (2015). Small-world networks in professional football: conceptual model and data. *Eur. J. Hum. Mov.* 35, 85–113.

Gama, J., Passos, P., Davids, K., Relvas, H., Ribeiro, J., Vaz, V., et al. (2014). Network analysis and intra-team activity in attacking phases of professional football. *Int. J. Perform. Anal. Sport* 14, 692–708. doi: 10.1080/24748668.2014.11868752

Grund, T. U. (2012). Network structure and team performance: the case of english premier league soccer teams. *Soc. Netw.* 34, 682–690. doi: 10.1016/j.socnet.2012.08.004

Hopkins, W. G. (2017). A spreadsheet for deriving a confidence interval, mechanistic inference and clinical inference from a P Value. *Sportscience* 10, 16–20.

Korte, F., and Lames, M. (2018). Characterizing different team sports using network analysis. *Curr. Issues Sport Sci.* 3:005. doi: 10.15203/CISS_2018.005

Linke, D., Lames, M., and Link, D. (2019). *Football-Specific Validity and Reliability of ChyronHego's TRACAB Player Tracking Technologies. Final Internal Report*. Munich: Technical University Munich.

Mclean, S., Salmon, P. M., Gorman, A. D., Naughton, M., and Solomon, C. (2017). Do inter-continental playing styles exist? Using social network analysis to compare goals from the 2016 EURO and COPA football tournaments knock-out stages. *Theor. Issues Ergon. Sci.* 18, 370–383. doi: 10.1080/1463922x.2017.1290158

Mclean, S., Salmon, P. M., Gorman, A. D., Stevens, N. J., and Solomon, C. (2018). A social network analysis of the goal scoring passing networks of the 2016 European football championships. *Hum. Mov. Sci.* 57, 400–408. doi: 10.1016/j.humov.2017.10.001

Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Soc. Netw.* 27, 39–54. doi: 10.1016/j.socnet.2004.11.009

Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., and Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. *J. Sci. Med. Sport* 14, 170–176. doi: 10.1016/j.jsams.2010.10.459

Pena, J. L., and Touchette, H. (2012). A network theory analysis of football strategies. *ArXiv* 1206.6904. doi: 10.1016/j.humov.2017.08.022

Pina, T. J., Paulo, A., and Araújo, D. (2017). Network characteristics of successful performance in association football. A study on the UEFA Champions League. *Front. Psychol.* 8:1173. doi: 10.3389/fpsyg.2017.01173

Praça, G. M., Lima, B. B., Bredt, S. D. G. T., Clemente, F. M., and Andrade, A. G. (2019). Influence of match status on players' prominence and teams' network properties during 2018 FIFA World Cup knockout phase. *Front. Psychol.* 10:695. doi: 10.3389/fpsyg.2019.00695

Ramos, J., Lopes, R. J., and Araújo, D. (2018). What's next in complex networks? Capturing the concept of attacking play in invasive team sports. *Sports Med.* 48, 17–28. doi: 10.1007/s40279-017-0786-z

Ribeiro, J., Silva, P., Duarte, R., Davids, K., and Garganta, J. (2017). Team sports performance analysed through the lens of social network theory: implications for research and practice. *Sports Med.* 47, 1689–1696. doi: 10.1007/s40279-017-0695-1

Sarmento, H., Clemente, F. M., Araújo, D., Davids, K., McRobert, A., and Figueiredo, A. (2018). What performance analysts need to know about research trends in association football (2012–2016): a systematic review. *Sports Med.* 48, 799–836. doi: 10.1007/s40279-017-0836-6

Tenga, A., Ronglan, L. T., and Bahr, R. (2010). Measuring the effectiveness of offensive match-play in professional soccer. *Eur. J. Sport Sci.* 10, 269–277. doi: 10.1080/17461390903515170

Trequattrini, R., Lombardi, R., and Battista, M. (2015). Network analysis and football team performance: a first application. *Team Perform. Manag.* 21, 85–110. doi: 10.1108/tpm-03-2014-0016

Vilar, L., Araújo, D., Davids, K., and Bar-Yam, Y. (2013). Science of winning soccer: emergent pattern-forming dynamics in association football. *J. Syst. Sci. Compl.* 26, 73–84. doi: 10.1007/s11424-013-2286-z

Yamamoto, Y., and Yokoyama, K. (2011). Common and unique network dynamics in football games. *PLoS One* 6:e29638. doi: 10.1371/journal.pone.0029638

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.